DOCUMENT RESUME

ED 105 891 IR 001 929

AUTHOR	Martin, Thomas H.
TITLE	A Feature Analysis of Interactive Retrieval Systems. Final Report.
INSTITUTION	Stanford Univ., Calif. Inst. for Communication Research.
SPONS AGENCY	National Science Foundation, Washington, D.C. Office of Science Information Services.
REPORT NO	SU-COMM-ICR-74-1
PUB DATE	Sep 74
NOTE	105p.; Report period October 1972 through October 1974
EDRS PRICE	MF-$0.76 HC-$5.70 PLUS POSTAGE
DESCRIPTORS	*Comparative Analysis; Computer Programs; Computer Storage Devices; Design; Human Engineering; *Information Retrieval; *Information Systems; Interaction; Management Systems; *Man Machine Systems; On Line Systems; *State of the Art Reviews
IDENTIFIERS	*Interactive Retrieval Systems

ABSTRACT

The command language features of 11 on-line information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal. Functional areas considered are: becoming familiar with the system, receiving help when in trouble, regulating usage, selecting a data base, formulating simple queries, expressing single concepts, interconnecting concepts, displaying results simply, and controlling the display. Features felt most essential to on-line searching are live help, users' guides, boolean operators, search field control, suffix removal, relational operators, dictionary access, request sets, search review, pre-defined formats, on-line formatting, and off-line printing. (SK)

ED105891

a report of the

INSTITUTE FOR COMMUNICATION RESEARCH
STANFORD UNIVERSITY

IR 001 929

ED105891

Report SU-COMM-ICR-74-1

A FEATURE ANALYSIS OF INTERACTIVE RETRIEVAL SYSTEMS

September 1974

Thomas H. Martin

Institute for Communication Research
Stanford University
Stanford, California 94305

Final Report for Period October 1972 - October 1974

Prepared for:

Office of Science Information Services
National Science Foundation
Washington, D.C. 20550

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Research supported by the National Science Foundation under Research Grant NSF GN 36160 awarded to Stanford University in October 1972

Principal Investigator: Edwin B. Parker, Ph.D.

Findings, opinions, and conclusions are those of the author or the principal investigator or both, and do not necessarily represent the views of the National Science Foundation

BIBLIOGRAPHIC DATA SHEET	1. Report No. SU-COMM-ICR-74-1	2.	3. Recipient's Accession No.
4. Title and Subtitle A FEATURE ANALYSIS OF INTERACTIVE RETRIEVAL SYSTEMS			5. Report Date SEPTEMBER 74 6.
7. Author(s) THOMAS H. MARTIN			8. Performing Organization Rept. No.
9. Performing Organization Name and Address INSTITUTE FOR COMMUNICATION RESEARCH Stanford University Stanford, California 94305			10. Project/Task/Work Unit No. 11. ~~Contract~~/Grant No. NSF GN 36160
12. Sponsoring Organization Name and Address OFFICE OF SCIENCE INFORMATION SERVICES National Science Foundation Washington, D.C. 20550			13. Type of Report & Period Covered Final: 10/72-10/74 14.

15. Supplementary Notes

NONE

16. Abstracts

The command language features of eleven different on-line information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal. Functional areas considered are: becoming familiar with the system, receiving help when in trouble, regulating usage, selecting a data base, formulating simple queries, expressing single concepts, interconnecting concepts, displaying results simply, and controlling the display. Features felt most essential to on-line searching are live help, users' guides, boolean operators, search field control, suffix removal, relational operators, dictionary access, request sets, search review, predefined formats, on-line formatting, and off-line printing. It is concluded that no sharp distinction exists between management information and bibliographic retrieval. The report is intended for use by designers of interactive retrieval systems and by students of system design

17. Key Words and Document Analysis. 17a. Descriptors

*Information retrieval
*Man-machine systems
*Design
*Information systems
Data storage systems
Psychometrics
Human factors engineering

17b. Identifiers/~~Open-Ended Terms~~

*Man computer interfaces
*Interactive retrieval systems
*Man computer communication
Management information systems
On-line systems
File management systems
Data management systems
Bibliographic information
User on-line interaction
On-line information retrieval
Human interaction with computers

17c. COSATI Field/Group 5B, 9B, 5A, 5E, 62B

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
No restrictions on distribution; Available from National Technical Information Service, Springfield, VA 22151	UNCLASSIFIED	103
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (REV. 3-72) THIS FORM MAY BE REPRODUCED USCOMM-DC 14952-P72

TABLE OF CONTENTS

1.0 Introduction
- 1.1 The Intended Audience 1
- 1.2 The Feature Analysis 2
- 1.3 How to Read the Report 5
- 1.4 Related Work 6

2.0 The Systems Selected
- 2.1 Introduction 11
- 2.2 Historical Perspective 13
- 2.3 The System's Environment 18

3.0 Instructional, Diagnostic, and Control Features
- 3.1 Introduction 24
- 3.2 Becoming Familiar with the System 25
- 3.3 Receiving Help when in Trouble 29
- 3.4 Regulating Usage 32

4.0 Query Formulation Features
- 4.1 Introduction 35
- 4.2 Data Base Selection 36
- 4.3 Formulating Simple Queries 40
- 4.4 Expressing a Single Concept 46
- 4.5 Interconnecting Concepts 56

5.0 Result Manipulation Features
- 5.1 Introduction 63
- 5.2 Displaying Results Simply 64
- 5.3 Controlling the Display 67

6.0 Conclusions
- 6.1 Introduction 75
- 6.2 Minimal Features 76
- 6.3 Principles for User-Oriented System Design . 81
- 6.4 A Future for Interactive Retrieval 83

7.0 Appendix 86

1.0 INTRODUCTION

1.1 THE INTENDED AUDIENCE

This report is intended for use by designers of interactive retrieval systems and by students of system design. In the report command language features of eleven different online information retrieval systems are presented in terms of the functional needs of a searcher sitting at a terminal.

There are a number of assumptions underlying the report that should be made clear. The first is that there has been enough time for competition in the marketplace since initial implementation for systems that do not respond to user needs to disappear. The second assumption is that command languages are sufficiently easy to change so that reoccurring suggestions from users are reflected in revised versions of systems. A third is that most (but not all) of the systems considered in the report were developed independently of the others. One can see this by noting that systems vary widely in the words used for invoking functions while they vary only slightly in the types of information users must supply when invoking the functions. A fourth assumption is that in the absence of totally new technologies, the functional needs of the searcher are not likely to change significantly. The net effect of the argument is that the juxtaposition of features from relatively successful operational retrieval systems can tell one a great deal about the needs of the interactive searcher. Hopefully, the report will ease the transition to a common vocabulary for talking about searching, and perhaps facilitate the development of a common interface for invoking different systems. At the very least, it should reduce chances that future designers build new systems in ignorance of experience acquired by their predecessors.

There are other audiences that are likely to think the report responds to their needs. Purchasers of interactive retrieval system service and searchers themselves probably will assume that the report can help them choose between systems. To the extent that it suggests what to look for, it is useful, but it should not be relied upon for more. First of all, the system descriptions are intentionally out of date. Secondly, there are many factors a purchaser should consider that are not discussed here - things like cost, reputation for service, and ability to tailor a contract to the specific needs of a purchaser. Without information about the particular set of data bases to be accessed, the educational background of users, and the usage loads at various times of the day and year, no adequate rating of systems is possible. Purchasers of systems or service should plan to carry out benchmark studies in which they determine what a typical load would look like and would cost if run on each of the systems under consideration. Finally, it should be pointed out that many more systems are available in the marketplace than are considered in this report. The National Bureau of Standards (3) has recently put together an index of interactive information systems.

1.2 THE FEATURE ANALYSIS

In order to assess the value of the report and the reasoning behind the selection of materials, the reader must understand the procedures followed in compiling it. The major reason for carrying out the analysis was to bring system designers together and to focus their attention upon similarities and differences in existing user interfaces. Interpersonal discussion is one of the most effective methods for changing human behavior, and user interfaces will not move toward commonality without first moving system designers toward commonality.

User interfaces are not the major interest of most designers. Designers spend most of their time making sure that their systems are reliable and cost effective. What they would most like to know and are least willing to reveal are strategies for increasing system throughput without degrading performance, or strategies for attracting and retaining a money-making mix of users. We intentionally did not probe too far into system internals or system clientele because these were felt to be privileged information. The user interface was a neutral topic for designers yet of great interest to searchers.

A three-day workshop was held at Stanford University April 23-25, 1973. Representatives of each of the eleven systems, a five-man panel of experts, and a handful of invited guests met for intensive discussions and system demonstrations. The five two-hour discussions focused on 1) the searcher/task environment, 2) the data base environment, 3) & 4) search and display features necessary for information retrieval, and 5) instructional and diagnostic features. At least one forty-minute demonstration was given of every system, and most systems were displayed twice. Video tapes were made of six of the system demonstrations directly off of the terminal and are available on loan from ASIS's SIG/UOI (the user on-line interaction special interest group of the American Society for Information Science).

In prepartion for the workshop, user manuals were solicited from each of the system representatives and each representative was visited for half a day. The manuals were used to extract much of the material that appears in the report, and the site visits were used to fill in missing gaps and to probe representatives regarding the user population. It rapidly became clear that designers knew little about the habits and characteristics of individual users. In most cases clients were known only by an account number

that lumped together a number of different searchers. Only gross generalizations were available regarding whether end users were carrying out their own searches, whether searchers were frequently switching between data bases, or whether they tended to use video terminals instead of teletypes. After the site visits and before the workshop, matrices were drawn up summarizing the degree to which each of the eleven systems incorporated various features. At the workshop the matrices were used to structure discussion between designers. Revised versions of the matrices have been included as an appendix to the report. Revisions have been made partially at the request of designers and partially to make the matrices an accurate summary of the total report. System representatives have also reviewed the chapters of the report dealing with their systems and have made revisions so that the report accurately states the status of their systems as of April, 1973.

We would like to thank Donald Black and Robert Katter from System Development Corporation, Mark Radwin and Roger Summit from Lockheed, Lawrence Stevens and Howard Coleman from Informatics, Stanley Friedman from IBM, David Colombo and John Fried from Battelle Memorial Laboratories, Richard Giering from Mead Technology Laboratories, Donald Hillman and Louis Stern from Lehigh University, Benjamin Mittman and Wayne Dominick from Northwestern University, Richard Marcus from MIT, and Charles Goldstein from the NASA Lewis Research Center for their contributions to the workshop and report.

In addition to the written report, a twenty-minute color 16mm. film called "Access" has been prepared for introducing college level students to interactive searching. It is being distributed by the Extension Media Center, University of California at Berkeley. In the film, viewers are informed that searching is like looking for a needle in a haystack and that it

involves the entering of a request, revision, supplementing with related terms, browsing, and display of results. They are shown a number of different types of users searching for different kinds of information in different kinds of data bases. Information technology is presented in an historical persepective as a means whereby the special interests of small groups can be served, and will be served more effectively in the future by means of networks, inexpensive storage media, and by human information brokers. The film can be used in conjuction with the report as an introduction to interactive searching.

1.3 HOW TO READ THE REPORT

The report itself is divided into chapters dealing with the various functional needs of interactive searchers. Matters outside the control of retrieval system designers are not included. For example, little is said about interactive terminal features or about operating system login procedures. Issues that are beyond the concern of searchers are not included. No mention is made, for instance, of file structures, parsing techniques, or updating techniques. There are a number of issues that it would have been nice to include but which were not since data was not available. For example, nothing is said about average and worst case delays due to system failure, or about recall/precision measures of system effectiveness.

The appendix to the report acts as an index to the subsections. The rows in the four matrices direct the reader to the related subsections. Subsections generally consist of statements supported by command language syntax. Rather than adhering to a uniform notation for commands, presentation techniques are tailored to the special circumstances of each subsection so that the supporting material will be as readable as possible. System names have been pulled into the margin so that they either can be ignored

or so that one can locate the manner in which a particular system has implemented a particular feature. Chapters begin with tutorial material placing the functional need in context and contrasting its use in interactive searching with its role in other types of systems. Then features which are encountered by default are introduced followed by features that must be specifically invoked. In some chapters features are organized in terms of subfunctions rather than in terms of how they are invoked. The report concludes with general recommendations to the designers of future systems.

1.4 RELATED WORK

The report fits into a background of activity that has been taking place over a number of years. In January, 1971 a workshop dealing with the user interface for interactive bibliographic searching (15) was sponsored by the American Federation of Information Processing Societies (AFIPS). The thirty-eight participants at the workshop represented a number of different interactive information retrieval systems. Many participants were unfamiliar with each other's systems and many expressed a desire to compare systems so that a basis could be laid for further discussion. Available comparisons of systems by Seidon (14) and Welch (16) were effectively out of date. The participants drew up a list of the functional areas involved in searching that form the basis for chapter divisions in the current report.

One of the participants at the 1971 workshop was William Olle, who headed a similar effort to place features in a functional framework for data base management systems (2) under the auspices of the Conference on Data Systems Languages (CODASYL). While there is an overlap between interactive retrieval langauages and data base languages, Olle indicated that he thought a similar feature analysis should be conducted for information retrieval systems (12). When the designers of the eleven retrieval systems

met at the April, 1973 Stanford workshop, one of their conclusions was that no sharp disctinction should be made between retrieval of management and bibliographic information.

The National Bureau of Standards has for a number of years kept track of the state of the art in information retrieval system development (3, 4, 6, 7). As the field has stabilized and pressure for standards has increased, NBS representatives have started laying a groundwork for uniform login and logout procedures.

While the National Bureau of Standards representatives are not attempting to impose standards in functional areas of searching, Reintjes and Marcus (13) at M.I.T. have been attempting to develop a translating interface language that can be used for coupling systems so that a searcher need not worry about which retrieval system he is using. Although they have made some progress the task is vast and much remains to be done.

At the other end of the spectrum, a survey under the direction of Carlos Cuadra at System Development Corporation is underway in an attempt to discover what purchasers, intermediaries, and end users actually think about the systems they have been using. In Europe, D.I. Raitt of the European Space Research Organization is using the same questionnaire to discover how Europeans are reacting to the systems available to them.

There are a number of places one can turn to for more information about interactive searching. A thorough and excellent introduction to the field is Lancaster's Information Retrieval On-line (5). If one prefers collections of papers, Walker's Interactive Bibliographic Search: The User Computer Interface (15), Meadow and Henderson's Interactive Bibliographic Systems (11), or May's Automated Law Research (10) are recommended.

There are not as many places where one can turn for information regarding the design of user/computer interfaces. James Martin's Design of Man-Computer Dialogues (8) is one of the only textbooks. The Annual Review of Information Science and Technology has recently included chapters on "The User Interface in Interactive Systems" (1, 9). The Special Interest Group "User On-line Interaction" of the American Society for Information Science periodically holds technical sessions dealing with user interface problems.

1. Bennett, John L. "The User Interface in Interactive Systems". In Cuadra, C. (Ed.) Annual Review of Information Science and Technology, Vol. 7. Washington, D.C.: American Society for Information Science, 1972, pp. 159-196.

2. CODASYL Systems Committee Technical Report. Feature Analysis of Generalized Data Base Management Systems. New York, N.Y.: Association for Computing Machinery, May, 1971, 520pp.

3. Fife, Dennis W. et al. "A Technical Index of Interactive Information Systems". Washington, D.C.: National Bureau of Standards, February 1974, 68p. (NBS Technical Note 819).

4. Fong, Elizabeth. "A Survey of Selected Document Processing Systems". Washington, D.C.: National Bureau of Standards, October 1971, 65pp. (NBS Technical Note 599).

5. Lancaster, F. Wilfred. Information Retrieval On-line. Los Angeles, Ca.: Melville Publishing Co., 1973, 596pp.

6. Marron, Beatrice. "A Study of Six University-Based Information Systems". Washington, D.C.: National Bureau of Standards, June, 1973, 95pp. (NBS Technical Note 781).

7. Marron, Beatrice et al. "A Mechanized Information Services Catalog". Washington, D.C.: National Bureau of Standards, February, 1974, 47pp. (NBS Technical Note 814)

8. Martin, James. Design of Man-Computer Dialogues, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1973, 559pp.

9. Martin, Thomas H. "The User Interface in Interactive Systems". In Cuadra, C. (Ed.) Annual Review of Information Science and Technology, Vol. 8. Washington, D.C.: American Society for Infromation Science, 1973, pp. 203-219.

10. May, Ronald A. (Ed.) Automated Law Research. Chicago, Ill.: American Bar Association, 1973, 182pp.

11. Meadow, Charles T. and Henderson, Madeline B. (Eds). Interactive Bibliographic Systems. Oak Ridge, Tenn.: U.S. Atomic Energy Commission, April 1973, 205pp. (CONF-711010).

12. Olle, T. William. "A Comparison between Generalized Data Base Management Systems and Interactive Bibliographic Systems". In Walker, D. (Ed.) Interactive Bibliograph Search: The User/Computer Interface, Montvale, N.J.: AFIPS Press, 1971, pp. 203-213.

13. Reintjes, J. Francis and Marcus, Richard S. "Interim Report on Research in the Coupling of Interactive Information Systems". Boston, Mass.: Massachusetts Institute of Technology, Feb., 1974, 36pp.

14. Seidon, H. R. "A Comparative Analysis of Interactive Storage and Retrieval Systems." Santa Monica, Ca.: System Development Corporation, Jan., 1970, (Report #TM-4421).

15. Walker, Donald E. (Ed.) Interactive Bibliographic Search: The User/ Computer Interface. Montvale, N.J.: AFIPS Press, 1971, 311pp.

16. Welch, Noreen O. "A Survey of Five On-line Retrieval Systems". Washington, D.C.: Mitre Corporation August, 1968, (Report #MTP-322).

2.0 THE SYSTEMS SELECTED

2.1 INTRODUCTION

Eleven systems (Battelle's BASIS, IBM's STAIRS, Lockheed's DIALOG, MIT's INTREX, NASA Lewis Research Center's NASIS, Lehigh's LEADER, Mead Technology Laboratories' DATA CENTRAL, Informatics' RECON, System Development Corporation's ORBIT II, Northwestern's RIQS, and Stanford's SPIRES II) were chosen for inclusion in the comparative analysis. The criteria used for selecting systems were that the system had to be 1) operational, 2) on-line and interactive, 3) able to handle multiple users simultaneously, 4) able to handle multiple data bases, 5) able to process data bases with variable length entries and elements, 6) demonstrable to the public, and 7) primarily oriented toward information storage and retrieval. In only one case was an exception made from these criteria: INTREX does not customarily handle more than one data base.

Many systems met the criteria of the comparative analysis but were not included. A number are close relatives of systems included in the analysis (ELHILL, NASA/RECON, BCN, LEXIS) so could be excluded without prejudicing the generality of the comparison. The versions included in the analysis were chosen because the person making the greatest contribution to the design of the system could represent it. Both DIALOG and RECON were included because Roger Summit of Lockheed and Larry Stevens (of NASA in 1968, of Informatics during the comparative analysis) worked together to establish the specifications for the original NASA/RECON.

Other systems were developed outside of the United States (QUIC/LAW of Canada, ISIS of Sweden) and it was felt that since system representatives

were being paid to come to the Stanford workshop, that it would be best to limit the analysis to U.S. systems.

Some systems came very close to being included but were not for extraneous reasons. Infodata's INQUIRE system was to be included but bowed out a month before the workshop because of the press of work. The New York Times' system would have been included except that it required a terminal that was not available on the West coast, was used with a single data base, and was written at IBM Federal Systems Division so would not have been represented by a system designer. Lawrence Livermore Laboratories' MASTER CONTROL would have been included except that it runs on a very unusual computer and can only be made publicly accessible by moving it to a declassified computer. The Defense Documentation Center's system was considered but not included for similar reasons. MRI's System 2000 would have been included except that it was discovered too late. Other systems considered but not used were: University of Washington's SOLAR, University of Pittsburgh's PIRETS, Syracuse's SUPARS, and Stanford Research Institute's Augmented Intellect.

No inference should be drawn that a system is in any way inadequate because it was not included in the analysis. We sought to limit the number of participating systems so that we could conduct a manageable workshop. Hopefully, the items analyzed throughout the report will prove to be sufficiently comprehensive so that other systems can be added on at a later date.

2.2 HISTORICAL PERSPECTIVE

Systems were selected to participate in the comparative analysis as of January, 1973. System representatives came together April 23-25, 1973 for the comparative analysis workshop. Throughout the rest of the report the present is assumed to be April 23, 1973, and this will be emphasized in the next two sections. However, the systems came into existence much earlier, and most have continued to exist beyond that date. In this section, each of the selected systems is introduced and presented as it has developed over time.

2.2.1 BASIS was developed at Battelle Memorial Institute's Columbus Laboratories and was first put into service during July, 1970. Initially intended for use by Battelle's information analysis centers and their clients, its use has broadened to cover many applications of information processing and modeling. While Battelle will sell or lease software, and will put up a customer's data base, potential clients are judged in terms of the research content of their proposed effort. BASIS

Since April, 1973 the system has also been installed in Japan, the largest bibliographic data base has been removed, a monitor has been developed, functions have been extended for interactive modeling and data analysis, and the system is being used for transportation and energy planning support.

2.2.2 STAIRS was developed by International Business Machines and was first put into service during July, 1972. While it was initially intended for use in-house (in the U. S. and in Europe), it is now being marketed as a software package. STAIRS

2.2.3 DIALOG was developed at Lockheed and was first put into service during 1966. The 1968 version of DIALOG was adapted by Lockheed to NASA's requirements (and was named RECON) and is now in the public domain. Since 1968 the system has been extended in many ways. While it is still possible to buy or lease the software, Lockheed encourages people to subscribe as interactive searchers. DIALOG

Since April, 1973 a great many more data bases have been put online, including the National Agricultural Library's CAIN (Cataloging and Indexing), the American Psychological Association's Psychological Abstracts, INFORM COMPENDEX, INSPEC, Chemical Abstract's CONDENSATES, and others.

2.2.4 INTREX was developed at the Massachusetts Institute of Technology and was first put into service during 1969. It was originally intended as an experimental system (but in an operational environment) for trying out new approaches to in-depth indexing and interaction via terminals which could display both alphanumeric characters and microfiche. INTREX

Since April, 1973 the system has been taken down and is being reprogrammed for use on an IBM 370/165 computer. In a related effort, an interface is being designed that translates what one requests using a single command language into the language of one or more other systems.

2.2.5 NASIS was developed at NASA's Lewis Research Center and was first put into service during 1970. The revised version of the system discussed in this report was in operation only a few months during 1972. NASIS was intended for use as a data base management system. Its major use was with descriptors of photographs of the United States taken by the ERTS satellite. The system is no longer being supported by NASA, although the software is still available. NASIS

2.2.6 LEADER (often referred to as LEADERMART when used in conjunction with Lehigh's Mart Library) was developed at Lehigh University and first put into service during March, 1971. It is intended primarily for searching of bibliographic or textual data bases by students at Lehigh, although service was for a time being sold to outside subscribers. LEADER is unique in its emphasis upon natural language processing and de-emphasis upon boolean operators. LEADER

Since April, 1973 the large bibliographic data bases have been removed from the system and reprogramming efforts are underway to make it more usable with textual (non-indexed) material.

2.2.7 DATA CENTRAL was developed at Mead Technology Laboratories and was first put into service during 1968. At various times before April, 1973 it was used for searching Psychological Abstracts, Epilepsy Abstracts, and the case law of the state of Ohio. DATA CENTRAL is intended for use as a data base management system. Either the software can be leased or data bases can be placed on the parent system. DATA CENTRAL

Since April, 1973 the system has been revised and is currently being reprogrammed to run on the DEC PDP-11 series of computers.

2.2.8 RECON was developed by Lockheed for NASA and is maintained by Informatics (in conjunction with their STIMS file maintenance package). Since 1968 the system has been extended in many different ways. RECON is installed not only at NASA, but also at the Department of Justice, the National Oceanographic and Atmospheric Administration, and other government agencies. Informatics supports RECON at these installations as well as selling service and making their computer available for storage of data bases. RECON

Since April, 1973 The Environmental Protection Agency's ENVIRON data bases and George Washington University's POPINFORM data base have put onto the system. The TOXICON data base has been removed and has become the TOXLINE data base on ELHILL.

2.2.9 ORBIT II was developed by System Development Corporation in conjunction with the development of ELHILL for the National Library of Medicine. An early version of ORBIT was first put into service during 1970, and ORBIT II became operational in 1972. SDC both sells or leases software and also puts up data bases for subscription searching on the parent system. ORBIT

Since April, 1973 a new version of the system, ORBIT III, has been developed. In addition, a great many more data bases have been put online, including the National Agricultural Library's CAIN, COMPENDEX, NTIS, INFORM and others. System improvements are being implemented on a continuing basis.

2.2.10 RIQS was developed by Northwestern University and was first put in service in September, 1969 (then called RIMS). Since that time it has undergone extensive changes and an instructional interface RIQSTUTOR has been built for it. RIQS is intended for maintenance and searching of small-to-medium size data bases whether they are bibliographic, textual, or numerical. It can be used in conjunction with graphics plotting and statistical analysis. RIQS

Since April, 1973 the RIQS User's Manual has been completely rewritten, a procedural capability has been developed to enhance both interactive searching and statistical data analysis, and an interactive graphics package has been developed. The system has been installed in Europe and is actively being marketed in the United States.

2.2.11 SPIRES-II was developed by Stanford in conjunction with the development of the BALLOTS library cataloging system. An early version of SPIRES was first put in service in 1969, and SPIRES II became operational during September, 1972. SPIRES II can handle a wide variety of different types of data bases, including those having hierarchies of fields within fields. While it is possible to transfer the software or to put a data base up for searching, SPIRES is intended primarily for searching and file maintenance by people at Stanford. SPIRES

Since April, 1973 the system has been revised substantially. It now provides for offline report generation and predefined formats online. As it is being modified to run on the IBM 370/158, it is being augmented to have host language capabilities.

2.3 THE SYSTEM'S ENVIRONMENT

Each of the systems must reside on a computer, under an operating system, (perhaps) connected by a data communications network to a terminal where a user searches in one or more data bases. The data base may belong to the user and is being stored on his computer with software leased from the system owner, the data base may be stored on the system owner's computter, or the user (as is customary) may have no ownership interest in the data base. In this section most of these aspects of the eleven systems are considered. Missing from what follows is a clear picture of the users, how frequently they search, and the problems they run into. System representatives were asked to characterize their users but few could do so.

Outside of the universities, most searchers are intermediaries rather than end users (but not all). An NSF-sponsored study is being conducted by Carlos Cuadra of System Development Corporation to find out the answers to user-related questions.

2.3.1 As of April 23, 1973 slightly modified versions of the system were being operated by others. These versions of the system would have qualified for inclusion in the comparative analysis.

a. National Library of Medicine's
   ELHILL (popularly known as MEDLINE)
b. European Space Research Organization's
   ESRO
   National Aeronautics and Space Administration's
   NASA/RECON
   Atomic Energy Commission's
   AEC/RECON
c. State University of New York's
   BCN (Biomedical Communication Network)
d. Mead Data Central's
   LEXIS

a. ORBIT
b. DIALOG
b. RECON
c. STAIRS
d. DATA CENTRAL

2.3.2 As of April 23, 1973 bibliographic data bases containing more than one hundred thousand records were available on the parent system for commercial searching.

a. ERIC (Educational Resources Information
   Center's data bases)
   CHEMON (Chemical Abstract's
   Chemical Condensates)
   MEDLINE (National Library of Medicine's
   MEDLARS)
b. ERIC
   PANDEX (CCM Information Services' data base)
   NTIS (National Technical Information Service's
   U.S. Government Research and Development
   Reports)
c. TOXICON (National Library of Medicine's
   Toxicology data base)
   ENVIRON (Environmental Protection Agency
   data base)

a. ORBIT
b. DIALOG
c. RECON
d. BASIS*
e. LEADER**

d. NTIS
Chemical Abstract's Condensates
* Science Information Associates handled marketing of searching service.

e. Chemical Abstract's Condensates
COMPENDEX (IEEE's Engineering Index)
** Metascience handled marketing of searching service.

2.3.3 In addition to the sites mentioned in 2.3.2, system software was also being used at the following locations.

a. ORBIT
b. RECON
c. STAIRS
d. DATA CENTRAL
e. NASIS

a. Department of State
Karolinska Institute (Stockholm, Sweden)
State University of New York (Syracuse)
b. Department of Justice (JURIS)
c. House of Representatives (Bill Status System)
d. Environmental Protection Agency
Wright-Paterson Air Force Base
Union Carbide
e. Carnegie-Mellon University

2.3.4 In addition to the data bases mentioned in 2.3.2, large data bases were available on the parent system for in-house use.

a. STAIRS
b. SPIRES
c. NASIS
d. INTREX

a. COMPENDEX
IBM's Technical Documents
b. MARC (Library of Congress's Machine Readable Cataloging data base)
c. ERTS (NASA's Earth Resources Satellite imagery data base)
d. the INTREX data base

2.3.5 As of April 23, 1973 the system was being used for some data bases that were neither bibliographic nor primarily textual.

RECON
DATA CENTRAL
BASIS
SPIRES
NASIS
RIQS

2.3.6 The system has capabilities equivalent to those of a host language. This means that programs written in COBOL or some other language can call upon the retrieval and display features.

DATA CENTRAL
NASIS

2.3.7 The system is operated in conjunction with a batch retrieval system. Only in the case of NASIS is it possible at the terminal to transfer a request from the online system to the batch system.

STAIRS
DIALOG
BASIS
LEADER
RIQS
NASIS
RECON

2.3.8 The major technique used by the system for translating requests into lists of records is to call upon previously constructed inverted indexes. Items in the index are kept in alphabetical order, and each item is a word, numeric value, or phrase followed by a list of pointers to records in the data base.

ALL
SYSTEMS
EXCEPT
RIQS

2.3.9 Pointers may indicate 1) the data base from which the record comes, 2) the field from which the item comes, 3) the word position within the field, 4) the item's importance to the record, 5) a date for the record, or 6) other information. Those systems that do not store data base and/or field information within the pointer keep separate inverted indexes for each searchable data base and/or field.

a. INTREX
b. DIALOG
c. STAIRS
d. SPIRES
e. LEADER
f. ORBIT
g. RECON
h. DATA CENTRAL
i. BASIS

a. 2,3,4 and 6
b. 1,2,3,4,5,6
c. 2,3, and 6
d. 5 and 6
e. 4 and 6
f. 4
g. 1,2,3, and 5
h. 2 and 3
i. 6 (for links)

2.3.10 An additional technique used by the system for translating requests into lists of records is to call upon every record in the current list (or in the data base if there is not a current list) and scan it to see whether it satisfied the query.

RIQS
INTREX
STAIRS
ORBIT
NASIS

2.3.11 As of April 23, 1973 the system was accessible through the Tymshare network (a commercial, leased telephone line, data communication network).

DIALOG
RECON
ORBIT
BASIS

2.3.12 Whether or not on the Tymshare network, a number of users were accessing the system from remote locations.

a. IBM has its own international data communications network
b. the federal telephone network (WATS)
c. WATS and privately leased links
d. NSF-sponsored leased line networks
e. Lockheed has its own high speed (480 cps) leased line network

a. STAIRS
b. NASIS
c. DATA CENTRAL
d. SPIRES
e. DIALOG

2.3.13 The user can search using 1) any of a wide range of different ASCII video and teletype terminals (like the Hazeltyne 2000 or the Texas Instruments Silent 700), 2) EBCDIC video and typewriter terminals (like the IBM 2260, IBM 3270 or IBM 2741, 3) graphics terminals (like the IMLAC or ARDS terminals).

a. 1 and 2
b. 2 and 3
c. 1,2 and 3
d. 1

a. ORBIT
a. DIALOG
a. RECON
a. STAIRS
a. DATA CENTRAL
a. SPIRES
a. NASIS
b. INTREX
c. BASIS
c. RIQS
d. LEADER

2.3.14 The system runs on a) the medium to large size IBM 360 and 370 series computers (like the IBM 360/40 or 370/145), b) the IBM 360/67, c) the CDC 6400, or d) the UNIVAC 1108. (INTREX ran on an IBM 7094.)

a. DATA CENTRAL
a. RECON
a. DIALOG
a. ORBIT
a. STAIRS
b. NASIS
b. SPIRES
c. RIQS
c. BASIS
c. LEADER
d. BASIS

2.3.15 The software is written primarily in a) FORTRAN, b) PL/1, c) Assembly Language, (SPIRES is written in PL/360 which acts like Assembly language but looks like PL/1. INTREX was written in an extension of Algol.)

a. BASIS
a. RIQS
a. LEADER
b. ORBIT
b. NASIS
c. DIALOG
c. RECON
c. STAIRS
c. DATA CENTRAL

3.0 INSTRUCTIONAL, DIAGNOSTIC, AND CONTROL FEATURES

3.1 INTRODUCTION

Before one can search for information using an interactive retrieval system, one must have access to the system and be able to use it. It is tempting to focus upon command language capabilities, forgetting that people need instruction, make mistakes, and occasionally need retraining. In this section we shall focus upon supportive and controlling features, so that in later chapters we can disregard them. A number of the features are actually not a part of the command language interface, but are necessary for it to function properly. For example, logging onto the computer will be discussed, as well as the training of users by human instructors and trouble-shooting by consultants when command language problems arise.

There is a danger when dealing with computer systems of assuming that everything should be computerized. Many system designers reported that the tutorial material available at the terminal is rarely invoked. There are at least three possible reasons for this. Interactive searching is still quite expensive (15-120 dollars/hr) and users may feel that there are cheaper ways to learn. It could also be that the range of behaviors to be taught or difficulties to be anticipated is so great that enough tutorial material can not be presented within a reasonable amount of time. Third, it is possible that the tutorial material available has not been reworked to the point where it makes sense to users. Whatever the reasons, all of the eleven systems depend heavily upon human assistance when dealing with users.

Few designers still argue that anyone can come in off the street and after fifteen minutes of instruction be searching like a professional. There are too many things that need to be taught. If the user is not familiar with his terminal, he has to be taught how to operate it. This may include plugging telephone receivers into modems, setting duplex switches, transmitting commands, and backspacing. Not only will the user have to learn how the retrieval language works, he probably will also have to learn how to search. Most non-mathematical users initially find boolean operators to be unnatural. It is only after some experience that they realize how important it is to fill out concepts and to progressively combine them. However, the topic usually requiring the most instruction is how to use the data base. Data bases differ markedly in structure, ranging from Medline with its MESH hierarchical thesaurus to ERIC with multiple index terms for the same topic, to Compendex with its index phrases, to still others for which every word in the text is inverted for searching. Some data bases rely heavily upon roles, others upon linkages, and some rely upon relations. For searching to be really effective, the query must be adapted to the peculiarities of the data base. Most searchers today are information specialists who, through constant practice, are comfortable with the terminal, command language, and data bases.

3.2 BECOMING FAMILIAR WITH THE SYSTEM

There are at least three types of people that receive instruction regarding system usage. At the highest level are purchasers of the system software or of storage space for a data base to be accessed using the system. These people generally receive extensive personalized instruction and will not be mentioned again. At the next level are individuals or institutions

for whom a searching account is to be opened. Finally there are individuals who search because they know someone who has access to a system. Commercial system representatives are most familiar with intermediate level people. Before opening an account, they make certain that the account holder or his representative receives instruction in system usage. University system representatives are generally more familiar with the lowest level users since system representatives and users are in close physical proximity. In both cases, it rarely happens that a person with an information need who has never searched using an interactive retrieval system before, sits down at a terminal by himself and begins searching. Almost always there is someone he goes to who knows how to use the system, and who will help him as he learns how to search.

3.2.1 The system customarily is invoked without going through special procedures known only to those who have been cleared for searching.

a. INTREX
b. SPIRES
b. BASIS

a. people generally go to a termainal that was logged on ahead of time by the INTREX staff.
b. people having account numbers for computer use in general can invoke the system by following normal logon procedures and typing SPIRES or BASIS.

3.2.2 Before users are cleared for searching, they are customarily required to attend a short course (usually about three days long).

ORBIT
DIALOG
STAIRS
NASIS
RECON
DATA CENTRAL
SPIRES

3.2.3 When users are told how they can invoke the system, they are customarily given personalized instruction for a few minutes.

LEADER
INTREX
RIQS
BASIS

3.2.4 As a part of their training, users are strongly encouraged to read the users' guide.

ORBIT
DIALOG
RECON
BASIS
SPIRES
NASIS
RIQS
DATA CENTRAL

3.2.5 Users are provided with pocket-sized folders (3"x8") summarizing the available commands for the system and, in at least two cases (*), the names of fields in one of the data bases.

STAIRS
DATA CENTRAL
SPIRES
INTREX*
DIALOG*

3.2.6 Even though the user has properly invoked the system, an attempt is made to catch the untrained user and provide him with instruction until he overrides it.

a. INTREX
b. SPIRES
c. ORBIT
d. BASIS
e. DATA CENTRAL
e. LEADER
f. STAIRS
g. RIQS
h. DIALOG

a. After the user types his name and address, the system responds "Welcome to Intrex _____. If you already know how to use Intrex, you may go ahead and type in commands. (Remember, each command ends in a carriage return.) Otherwise, for information on how to make simple searches of the catalog, type
info 2
or, to see the Table of Contents (Part I) of the Intrex Guide which will direct you to other parts of the Guide explaining how to make more detailed searches, type
info 1"

b. After the user types SPIRES, the system responds
"WELCOME TO SPIRES II, _____.
followed by messages of the day, then
IF IN TROUBLE TYPE HELP"

c. After the user is logged in, he is told which data base he is connected to. Then he is asked whether he wishes to have the new or experienced user format. If he responds that he wants the new user format, he is advised
"TYPE ONLY AFTER THE CUE 'USER:', WAIT AFTER THE CUE 'PROG:', ENTER SEARCH STATEMENTS ... DO YOU WISH A FURTHER SUMMARY OF OPERATING FEATURES? TYPE YES OR NO AND CARRIAGE RETURN."

d. After the user logs on he is asked
DO YOU DESIRE OPERATING INSTRUCTIONS?
TYPE YES OR NO
e. From the moment the user logs onto the system, he is advised regarding what to do next. If he has received training, he is able to enter commands not included in the advice.
f. After signing onto a data base, the user is informed that he can invoke the HELP function.
g. If accessed from RIQSTUTOR, the RIQS user is informed
"IF AT ANY TIME WHILE ENTERING SEARCH COMMANDS YOU NEED ASSISTANCE, TYPE HELP..."
h. As a part of the log on greeting, the user is given a telephone number for problems or assistance.

3.2.7 The knowledgeable person can receive enough instruction at the terminal in order to learn how to search. This may include 1) terminal and typing problems, 2) the command repertoire, 3) characteristics of the data base, 4) common pitfalls and their remedies, 5) recent revisions to the system, 6) sample searches to be emulated, or 7) hints for effective searching.

a. DATA CENTRAL
b. SPIRES
c. RIQS
d. INTREX
e. ORBIT
f. STAIRS
g. DIALOG

a. depending upon the data base 1,2,3,4,6 and 7
b. 2,3,5,6
c. 2,3,4,5,6,7
d. 1,2,3,5,6,7
e. 1,2,3,5
f. 2,3,6
g. 2,3,5

3.2.8 The user either can access training material passively via a tutorial or can actively use an index to call up pages of instruction.

a. DATA CENTRAL
b. SPIRES
c. RIQS
d. INTREX
e. ORBIT
f. STAIRS

a. WHAT# allows one to access pages of text. Many of these pages suggest other pages to access. There is a directory for finding out what text is associated with each number.
b. HELP instructs the novice to type EXPLAIN EVERYTHING (a directory of what is available). EXPLAIN EVERYTHING indicates that the user seeking instruction should type EXPLAIN SPIRES for tutorials. SHOW NEWS explains the latest changes in the system.

c. The user must attach RIQSTUTOR, which in turn attaches and communicates with RIQSONLINE. During the tutorial, control can be passed to RIQS via the EXIT command. Control can be returned to the tutor via the HELP command.
d. INFO# allows one to access pages of text.
INFO 2 acts as a directory.
NEWS allows one to find the latest changes in the system.
e. The new user receives tutorial information.
EXPLAIN EXPLAIN acts as a directory.
f. HELP allows one to access pages of tutorial explaining different aspects of the system.

3.3 RECEIVING HELP WHEN IN TROUBLE

It is a rare person who never makes a mistake. The range of problems that can arise when searching is enormous. At best the user realizes he has forgotten how to enter a command. Often he cannot figure out what a message from the computer means, but is satisfied when he receives a more verbose explanation. Problems become more difficult when the user is certain that one thing should have happened but instead another did. For example, the user asks for all records dealing with AMERICAN HISTORY OR UNITED STATES HISTORY and finds that the data base contains nothing. The problem could be that ORs are processed before ANDs, that the data base has been broken into individual words, and that implicit ANDs are inserted between words in a query. Thus the request effectively asks for records containing AMERICAN AND STATES AND HISTORY. Or the problem might be that the user returned the cursor before transmitting the line. Or the problem might be that the data base distinguishes between upper case and lower case letters, and that the upper case query did not retrieve any of the mixed case documents. While the above is hypothetical, it points out the range of causes there may be for a single problem.

Most of the systems do very little on-line to help users with their problems. Almost all of them provide users with telephone numbers to call as a last resort. They rely heavily upon good training in the first place, and interfaces that are so straightforward that users find it difficult to make errors.

3.3.1 The user who has switched to a terse form of interaction with the system, can switch back to a verbose form whenever he feels he needs reminding.

a. INTREX
b. ORBIT
c. DATA CENTRAL

a. LONG to get the verbose mode,
   SHORT to get back to the terse mode.
b. VERSION LONG to get the verbose mode,
   VERSION SHORT to get the intermediate,
   VERSION SYMBOLIC to get the terse mode.
c. When entering the data base, the option
   LONG or LKWIC is used to get the verbose
   mode, while SHORT or SKWIC is used to get
   the terse mode.

3.3.2 The user can ask for an explanation of the system prompt or response he has just received.

a. RECON
b. ORBIT
c. RIQS
d. SPIRES
e. STAIRS
f. DIALOG
g. NASIS
h. DATA CENTRAL

a. HELP error code
b. EXPLAIN will give an explanation of
   the latest prompt by the computer.
   EXPLAIN followed by the terse form of a
   system message will retrieve an
   explanation of the message.
   HELP will cause a menu of options to be
   displayed
   (no postings?, see results?, need
   to answer a question?, review features?,
   review data base?, command names?)
c. If the RIQSTUTOR is attached, then HELP
   invokes the tutor and it explains how the
   RIQS command component most recently used
   works.
d. EXPLAIN error code
   HELP will explain to the user the stage
   of searching he is currently in,
   and will suggest what he might do
   to proceed to the next stage.
e. ..HELP error code or ..HELP option
   ..HELP without a parameter prompts
   the user with a list of things
   for which he can receive explanations.

f. EXPLAIN error code
g. EXPLAIN error code
EXPLAIN RESPONSE, error code in order to find out what to do to correct an error.
EXPLAIN ORIGIN, error code in order to find out where an error code came from.
h. WHAT#
every prompt and response from the system is numbered so that fuller explanations can be retrieved.

3.3.3 The user can ask for an explanation of a specific command for which he remembers the name but probably not the format.

a. HELP command name
b. EXPLAIN command name
c. ..HELP command name
d. if the RIQSTUTOR is attached, then the user can ask for explanations of commands at any time during query formulation.
e. DATA CENTRAL has explanations of the specific commands in the WHAT file.

a. RECON
b. ORBIT
b. DIALOG
b. SPIRES
b. NASIS
c. STAIRS
d. RIQS
e. DATA CENTRAL

3.3.4 The user can have the processing of his search request traced so that he can get a better feeling for what went wrong.

a. ..SET DETAIL=ON causes the number of records and the number of occurrences for every term in the query to be listed, but does not show how the terms get combined.
b. COUNT causes the number of records remaining to be listed after each new term's stem is ANDed into the result.

a. STAIRS
b. INTREX

3.3.5 A portion of the user's guide is written in such a way that it helps the user with trouble-shooting.

BASIS
STAIRS
RECON
SPIRES
ORBIT
DATA CENTRAL
DIALOG

3.3.6 Possessors of account numbers are provided with a telephone number for calling up a human consultant who can help them with trouble-shooting. — ALL SYSTEMS

3.3.7 The user can interact in real time with the consultant via the command language. — a. NASIS, b. SPIRES, c. DIALOG

a. HELP followed by text prints the message at the consultant's terminal.
b. TO SPIRES takes the message composed using the text editor and puts it in the consultant's mailbox. Assuming that the user did not turn off monitoring, the consultant can look over the user's recent behavior from his consulting terminal.
c. SEND MESSAGE 1/ followed by text prints a message at the consultant's terminal, who then can send a message back or call the user.

3.4 REGULATING USAGE

There are a few features that users often do not perceive as beneficial, but that ultimately are for their benefit. Some features, like access restrictions, may rarely be brought to the attention of users. Others, like encouraging comments, may be highly visible. The features are included because they make it possible for system representatives to learn from the experiences of users and to regulate who has access to their systems.

3.4.1 When entering the system, the user must pass through the logon procedures of 1) a computer network, 2) an operating system, and 3) the retrieval system itself. He must know 4) an account number, 5) a terminal identifying code, and 6) a password.

a. 3 and 4; sometimes also 1,5, and 6
b. 2,6, and 3; sometimes also 1 and 5
c. 3 and 4; sometimes also 1 and 5
d. 3 and 4
e. 2,4,6 and 3; sometimes also 1,5,and 6
f. 2,4,5,6 and 3
g. 2,6 and 3
h. 2,4,6, and 3
i. 2,3,6, and sometimes 5
j. 3.4, and sometimes 6

a. ORBIT
b. BASIS
c. DIALOG
d. DATA CENTRAL
e. RECON
f. SPIRES
g. LEADER
h. RIQS
i. STAIRS
j. NASIS

3.4.2 By the time the user has logged on, his searching abilities have 1) been restricted to a subset of all possible data bases, 2) a subset of the fields in a data base's records, and 3) a subset of the records in the data base.

a. 1
b. 1 and 2
c. 1, 2, and occasionally 3

a. RIQS
a. BASIS
a. DIALOG
a. ORBIT
a. RECON
a. LEADER
b. DATA CENTRAL
c. NASIS
c. STAIRS
c. SPIRES

3.4.3 By the time the user has logged on, his command behavior is being monitored (disregarding purely charge monitoring) a) only as a contribution to aggregate statistics, b) on a command in core basis rarely used for tracking individual users, c) on an individual user basis so that users can both be studied and helped with problems, or d) on an individual basis but with an option for the user to turn monitoring off.

a. NASIS
b. ORBIT
b. RECON
b. DIALOG
c. LEADER
c. INTREX
d. SPIRES
d. RIQS

3.4.4 The user can enter comments as he searches that appear at the appropriate spot in the monitor log.

a. "COMMENT followed by text"
b. COMMENT prompts the user for lines of text until the user specifies STOP. Or
   * followed by text for notes to oneself.
c. COMMENT followed by text
d. The user is asked after completing use of RIQS to enter any suggestions or comments he might have.

a. ORBIT
b. LEADER
c. INTREX
d. RIQS

4.0 QUERY FORMULATION FEATURES

4.1 INTRODUCTION

Perhaps the major value of information retrieval systems is that one can easily gain access to small numbers of potentially relevant records that are buried in immense data bases. By trying out various criteria, by observing how the criteria pare away a data base, and by examining a scattering of retrieved records, the interactive searcher can revise his strategy until he senses that he can do no better. While a good searcher vacillates between expressing criteria and checking to see how well the criteria work, in this chapter we shall only discuss features for formulating queries.

Formulating queries for information retrieval systems is not the same as asking everyday questions. First of all, the context in which most everyday questions are asked greatly limits the possible responses. One does not ask a waitress how much it costs to fly to Tokyo. Secondly, most questions incorporate an understanding of how the respondent structures his knowledge. Thirdly, askers of difficult questions rarely expect thorough or precisely on-point answers. Users of information retrieval systems often do expect to receive thorough and on-point answers. They expect to get these from data bases that contain material contributed by many different people who structure their knowledge in different ways. Computers are not in eye-to-eye contact with humans and generally do not remember people from one session to the next. The information retrieval systems discussed in this report do not answer questions; they provide a "quick and dirty" means for narrowing down the search space so that the odds are improved that the user will find an answer to his question.

Nevertheless, formulating search queries has much in common with formulating questions. Words and phrases must be used that pinpoint the topic of discussion. Parts must be connected together in ways that suggest the desired relationships. Occasionally one needs to be able to emphasize the exactness of a phrase or to wave arms because a word is stuck on the tip of one's tongue. Just as most of us help others rephrase bad questions, interactive systems can prompt for clarification when confused, can lay out the possibilities where there are multiple interpretations, or can explain why it would cost too much to respond.

In discussing features, the reasons for having them will be considered along with alternate ways for implementing them. While an effort is made to discuss features without tying them to particular storage schemes or data bases, occasionally the discussion is made concrete. The similarity across systems comes not from a common understanding of human expression but from a common approach to storing data and an interest in bibliographic citations. As stated before, one of the reasons for this analysis is to assess how sensitive command language features are to current technologies. Where insensitive, perhaps functional needs can be formally recognized, and treated by all systems in a common manner.

4.2 DATA BASE SELECTION

In all of the eleven systems, one must select a data base before searching can begin. The word "data base" is used to refer not only to some logical file of similarly-structured records, but also to auxiliary files containing indexing terms, tutorials, or other material supporting the primary file. While it is generally true that the "data base" selected by the user corre-

sponds directly to the physically-stored version, there are exceptions (SPIRES, STAIRS, DATA CENTRAL) where the user-perceived data base may be a subset of the stored version. None of the systems allow the user at his terminal to patch together a conglomerate data base out of physically distinct ones. This means that a person interested in a topic spanning unconnected data bases must search each one separately. Only one system (LEADER) automatically carries the search request along when switching to a new data base. Other systems (DIALOG, STAIRS, NASIS, SPIRES) allow requests to be stored and called up for use with other data bases. In general, the more controlled the indexing vocabulary for a data base, the less likely queries composed for searching in it will transfer to other data bases. Even where a single query can be passed against multiple data bases, the user must combine the separate listings by hand. It remains for future studies to show how often users need to access more than one data base, and how bothersome it is that final results are not merged together automatically.

The features discussed below are those for selecting an initial data base and then for switching to a different one. Tutorial and other help-providing features are discussed in the previous chapter. Our conclusion there was that designers assume users do not need help with data base selection. Little or no advice is provided regarding subject coverage, indexing policies, or effective searching techniques. If systems are intended for use by those with information needs (e.g. scientists, doctors, lawyers, etc.), perhaps at least optional introductions should be provided for each of the available data bases.

4.2.1 The set of available data bases does not depend upon the user identification. — INTREX*

* INTREX makes available only a single data base. For use with other data bases, the system is called other names.

4.2.2 When the user identification happens to limit one to a single data base, that data base is attached automatically. — INTREX, ORBIT, RECON, DIALOG

4.2.3 When the user identification does not limit one to a single data base, one data base is still attached automatically. — RECON, ORBIT, DIALOG*

* DIALOG has a user-specified default data base.

4.2.4 The user is automatically presented with a list of the data bases he might attach. — NASIS, LEADER

4.2.5 The user can ask for a list of the data bases he might attach. — a. DIALOG, b. SPIRES, c. STAIRS, d. NASIS, e. ORBIT, f. RIQS, g. DATA CENTRAL

a. BEGIN or EXPLAIN FILES
b. SHOW SUBFILES
c. ..HELP BASES
d. FILES or RETRIEVE (where there is no default file name)
e. "FILES?"
f. BROWSE when using the RIQSTUTOR
g. WHAT 50

4.2.6 As a part of the available data base display, the user is asked to enter the identifying number of the one he wants to have attached. — LEADER, DIALOG, NASIS, RIQS*

* when using the RIQSTUTOR

4.2.7 The user, independently of an available data base display, indicates to the system which data base he wishes to have attached as his first data base.

a. DATA CENTRAL
b. RIQS
c. BASIS
d. STAIRS
e. SPIRES
f. DIALOG
g. NASIS

	SYSTEM PROMPT	RESPONSE
a.	ENTER FILE, MESSAGE OPTION	db, option
b.	?	ATTACH, logical unit, db
c.	ENTER NAME OF THE DATA BASE TO BE SEARCHED	db
d.	ENTER DATA BASE NAME	db password (password is optional)
e.	-?	SELECT db
f.	ENTER:	BEGIN BYPASS (for the user-specified default data base)
g.	-ENTER NASIS COMMAND:	RETRIEVE db

The following four items assume that the user wishes to switch to a new data base.

4.2.8 The user must first exit from the system, although he may still maintain contact with the time-sharing monitor.

INTREX
RIQS

4.2.9 The user indicates that he wants to be returned to the point just before he was asked to select his first data base.

a. STAIRS
b. LEADER
c. BASIS

a. ..CHANGE
b. CHANGE
c. RESTART

4.2.10 The user indicates to the system which data base to attach just as he did initially.

SPIRES
NASIS

4.2.11 A slightly different technique is used for switching to a new data base than was used initially.

a. RECON
b. DIALOG
c. DATA CENTRAL
d. ORBIT

a. BEGIN for a list of the attachable data bases and a prompt to select one, or BEGIN db# to bypass the listing and prompting
b. .FILE db#
c. FILE for a reminder of the current data base (and option), with a chance to switch by typing db, option
d. "FILE db"

4.3 FORMULATING SIMPLE QUERIES

Many people prefer to learn by doing, and often begin searching before they know how to. All of the eleven systems rely upon the user to formulate the query, but vary in how easy they make matters for the uninitiated. Since people often rely heavily upon first impressions, it is important to consider what each system presumes that the user knows about entering requests, how successfully the system re-orients him when he gets confused, and how rapidly it brings him to the point where he can display retrieved records. Since simplicity is gained at the expense of making presumptions, it is worthwhile to examine what assumptions each system makes, and how systems differ.

It is difficult to anticipate what an untrained user will consider to be a good request. Much depends upon the type of data base selected. In order to remove this variable, and since all of the eleven systems can be used with textual or bibliographic data bases, we shall assume that the user has selected a textual data base. According to designers, the great majority of requests are for content-related records. The untrained user is likely to rely upon prior experience with content-related searching. If he thinks about how he asks people questions, he probably will sketch out most aspects of his query, using good grammar and more words than absolutely necessary.

If he thinks about how he uses indexes or card catalogues, he probably will enter one or two words that capture a single aspect of the query. Instructors of interactive searching typically have to counteract defects in each of the analogies. The question-asker is encouraged to enter one concept at a time so that he can discover whether clarification is needed to fill out or delimit the concept. The index-thumber is encouraged to enter multiple concepts, connecting them together with boolean operators. Without instruction, the first individual may never retrieve anything, while the other may end up manually searching through voluminous output. The sample queries we will use are "the effects of television violence on children" for question-askers, and "television violence" for index-thumbers.

It should be pointed out that DIALOG and RECON attempt to discourage users from directly entering full queries, and instead encourage them to scan alphabetically adjacent index terms. They do this by having a row of keys on the terminal act like functions keys, and by having the one for EXPAND left of the one for SELECT. The untrained user who begins his request with EXPAND receives back a portion of the index, thereby increasing chances that he will not enter wordy requests in the future.

The reason for not discussing learning by doing in the chapter on instruction is that none of the systems treat simple searching as an opportunity for introducing the user to more sophisticated features. Many users never progress beyond simple searching. Only in the case of the LEADER system is simple searching likely to deal effectively with quite complex requests. LEADER attains this power by using the request as a device for transferring the searcher's attention to index phrase selection.

4.3.1 The system assumes the user wants to enter a search request after having selected the data base.

a. ENTER REQUEST
b. SS 1/C? -- SEARCH STATEMENT 1 OR COMMAND?
c. ENTER YOUR SEARCH ONE TERM AT A TIME
1/
d. PLEASE ENTER YOUR REQUEST. (END IT WITH A PERIOD)
e. TO FIND DOCUMENTS IN THE SYSTEM SPECIFY YOUR SEARCH REQUEST BY SUBJECT, AUTHOR, TITLE TERMS, OR COMBINATIONS OF THESE AS SHOWN IN THE 3 EXAMPLES BELOW:
SUBJECT XENON VISCOSITY (etc.)
f. ENTER SEARCH COMMAND OR TYPE HALT

a. DATA CENTRAL
b. ORBIT
c. BASIS
d. LEADER
e. INTREX
f. RIQS

4.3.2 The user must indicate to the system that he intends to enter a search request.

a. SELECT
b. ..SEARCH
c. FIND

a. NASIS
a. RECON
a. DIALOG
b. STAIRS
c. SPIRES

4.3.3 The user must specify the field to be searched.

a. SUBJECT
b. an index for the selected data base
c. RECORD (to search all fields)

a. INTREX
b. SPIRES
c. RIQS

4.3.4 The user must specify the logical relationship that the query value has in common with the field in question.

a. CONTAINS

a. RIQS

4.3.5 The user must specify the relationship holding between multiple words in the query value.

a. 'query words'
* this is true only for data bases not having multi-word index terms

a. RIQS
a. RECON*

4.3.6 The user must specify in the query what is to be done with the results of searching.

a. RIQS

a. DISPLAY field names or numbers or
   PLACE RECORDS IN SET set number

4.3.7 The end of the query must be designated in some manner (other than by a carriage return or transmit).

a. RIQS
b. LEADER

a. END
b. .

For the next seven items, the words making up the query take a form like THE EFFECTS OF TELEVISION VIOLENCE ON CHILDREN, where no record contains exactly this phrase nor does the user expect one to.

4.3.8 By expressing a request in this manner, the user is likely to receive worthwhile results.

a. INTREX
b. LEADER

a. While AND is the implied boolean connector, the effects of stemming and full indexing reduce the chances that the user will receive a null result.
b. The words are used to retrieve indexing phrases ranked from those containing all the words to those containing only one. The user then selects some of the phrases and these become his request.

4.3.9 The user will receive a null result with no suggestion for entering worthwhile requests.

DATA CENTRAL
DIALOG
RECON
NASIS
SPIRES*
ORBIT
RIQS

* for data bases where the implied boolean connector is an AND

4.3.10 The user will receive a null request but with a suggestion that improves his chances of entering worthwhile requests. — a. BASIS

a. NO SUCH TERM. WANT ADJACENT TERMS? YES:NO/

4.3.11 The user will receive a result, but neither is it very useful nor is he given suggestions for entering worthwhile requests. — STAIRS, SPIRES*

* for data bases where the implied boolean connector is an OR.

4.3.12 Common words like 'the', 'of', and 'on' are deleted from the query. — INTREX, LEADER, DATA CENTRAL*, SPIRES*, STAIRS*

* for those data bases where the data base manager has chosen this option

4.3.13 Words like 'effects' (but not 'children') are depluralized. — LEADER, INTREX, DATA CENTRAL*

* the data base manager specifies which depluralization rules he wants to incorporate

4.3.14 Words like 'television', 'violence', and 'children' have their affixes removed so that they become stems like 'televi', 'viol', and 'childr'. — INTREX

For the next five items, the words making up the query take a form like TELEVISION VIOLENCE where the user does not realize there is a difference between the exact phrase and the intersection of the two terms.

4.3.15	The request is interpreted as a search for the term TELEVISION VIOLENCE in the index. * for data bases not having multiple word index terms, the request will lead to a null result ** true only for data bases having multiple word index terms	BASIS ORBIT* DIALOG RECON** NASIS*
4.3.16	The request is interpreted as a search for phrases like TELEVISION AND VIOLENCE, TELEVISION VIOLENCE, TELEVISION OF VIOLENCE in the record.	DATA CENTRAL
4.3.17	The request is interpreted as a search for the exact phrase TELEVISION VIOLENCE in the record * true only for data bases with word proximity searching	RIQS RECON*
4.3.18	The request is interpreted as a search for documents indexed under (or containing) both the words TELEVISION and VIOLENCE. * true only for data bases with AND as the implied boolean connector	SPIRES* INTREX
4.3.19	The request is interpreted as a search for documents indexed (or containing) either the word TELEVISION or VIOLENCE. * true only for data bases with OR as the implied boolean connector	SPIRES* STAIRS
4.3.20	The request is interpreted as a search for documents indexed (or containing) either TELEVISION or VIOLENCE or both (whether or not the terms are contiguous). The results are ranked. * when using the RANK command	LEADER STAIRS*

4.4 EXPRESSING A SINGLE CONCEPT

The experienced searcher, unlike the novice, recognizes the importance of breaking queries into distinct concepts, and of polishing each concept until the right balance between precision and ambiguity has been achieved. In natural language communication, it is rarely necessary to focus upon isolated concepts or to worry about how they are expressed. The rich grammatical conventions that we use in natural language for embedding concepts are not available in command languages. By focusing upon concepts, logical bonds linking concepts become clearer, and so can be more easily reconstructed with boolean operators in a query. Concept linking will be discussed in the next section. The rich store of redundant expressions and interconnected ideas that recipients of messages possess is not built into most data bases. By attempting to thoroughly express each concept, new interpretations of the query come to light and shades of meaning stand out. In this section, features for refining concepts are discussed.

The most prevalent features are those for incorporating additional terms into the request. New terms can range from slight variations in spelling to semantic relatives. There are a number of reasons why so many features should be dedicated to incorporating new terms. Users, like most people, find recognition easier than recall. Alphabetical displays and displays of related terms stimulate the user with slight variants for his terms that he might otherwise forget to include. Bad spellers can use alphabetical displays in the same way they use dictionaries. Features that simplify entering large numbers of almost identical terms not only reduce the chances for error but also speed up the query formulation process.

Features for restricting the scope of terms are less prevalent. Most of the systems do not distinguish between upper and lower case letters in a query. Most avoid confusions about whether characters are part of a term or part of a command by eliminating special characters from index terms. (The techniques used by SPIRES and NASIS for handling special characters are not discussed in this report.) Many permit the user to search for exact phrases. A few make it possible for the user to limit how far apart words can appear in a retrieved document. All systems let the users restrict searching to one or more fields. Perhaps the restrictive power of the boolean AND (discussed in the next section) when used to combine different concepts is so great that the need for limiting features at the concept level is reduced.

The following thirteen items facilitate expanding the scope of concepts.

4.4.1 Various formats for expressing certain types of values can be recognized and automatically converted to a standard form before searching.

a. RIQS
a. DATA CENTRAL
b. SPIRES

a. 3 FT. 2 IN. converts into 38 IN.
   JUNE 16, 1973 converts into 06/16/73
b. FRED JONES converts into JONES, F.
   JUNE 1973 converts into 06/--/73.

4.4.2 The user can specify that values falling between a lower and upper bound should be incorporated into the query. An asterisk implies sequential searching. DATA CENTRAL and RIQS achieve this feature using relational operators and the AND logical operator.

a. NASIS
b. RECON
b. DIALOG
c. BASIS
d. ORBIT
e. SPIRES
f. STAIRS

a. TERM1 - TERM2 or TERM1: TERM2
b. TERM1: TERM2
   LIMIT set/## -## years only
c. (TERM1 - TERM2) numeric only
d. 19## THRU 19## years only
e. BETWEEN TERM1 and TERM2 numeric
   FROM TERM1 TO TERM2 only
*f. WL TERM1, TERM2 (within limits)
   OL TERM1, TERM2 (outside limits)

4.4.3 The user can specify criteria for numeric fields via relational operators. For all but RECON, the format is FIELD NAME OPERATOR TERM. For RECON the format is OPERATOR FIELD NAME TERM. STAIRS and RIQS are limited to sequential searching of numeric fields.

a. NASIS
b. RECON
c. SPIRES
d. RIQS
e. STAIRS
f. DATA CENTRAL

	>	≥	=	≠	≤	<
a.	>,GT	>=,GE,¬<	=,EQ,BT,><	¬=,NE	<=,LE,¬>	<,LT
b.	+	+=	=		-=	-
c.	>,AFTER	>=	=	¬=	<=	<,BEFORE
d.	GT	GE	EQ	NE	LE	LT
e.	GT	NL	EQ	NE	NG	LT
f.	>,GTR	NLS,/<	A=,AEQ	A/=,ANEQ	NGT,/>	<,LSS

4.4.4 Common prestored variants for a query term are automatically incorporated into the query (i.e. AEC also retrieves documents indexed under ATOMIC ENERGY COMMISSION).

STAIRS
DATA CENTRAL
BASIS

4.4.5 The user can specify that terms with the same root as the query term should be incorporated into the query (i.e. CHILD# acts like CHILD, CHILD'S, or CHILDREN).

a. DATA CENTRAL
b. STAIRS
c. SPIRES
c. DIALOG
d. LEADER
e. ORBIT
f. RECON
g. INTREX

a. TERM**** incorporates terms having (in this example) four or fewer characters following the root.
b. TERM $, TERM $n, MASK TERM, or MASK TERM n incorporate terms having an unlimited number or no more than n characters following the root.
c. TERM# for data bases defined to permit truncation (a different symbol could be designed as the truncation code in the data base definition).
d. TERM*
e. ALL TERM#
f. TERM:
g. TERM! (since truncation is the normal mode, this specifies an exact match on all characters)

4.4.6 The user can specify that terms matching the query mask should be incorporated into the query.

a. ON*LINE retrieves ON-LINE and ON LINE but not ONLINE
b. WITH ONLINE retrieves all of the above via something like sequential searching
c. ON#LINE retrieves ON-LINE and ON LINE but not ONLINE
d. STRINGSEARCH (FN) ON:LINE retrieves all of the above via sequential searching

a. DATA CENTRAL
b. SPIRES
c. DIALOG
c. ORBIT
d. ORBIT

4.4.7 The user can ask for a display of terms alphabetically adjacent to the one he designates. Possible features of a display are 1) that two or more of the displayed terms precede the term alphabetically, 2) that the terms are given identification so they can be referred to later, 3) that each term indicates how many records would be retrieved if it were entered in a search request, 4) that each term indicates how many terms are cross-indexed with it, and 5) that the term causing the display is highlighted or designated in some manner. (Parenthesis around FN means that specifying a field name or index is optional).

a. STAIRS
b. ORBIT
c. DIALOG
c. RECON
d. NASIS
e. BASIS
f. SPIRES
g. DATA CENTRAL

	COMMAND	RESPONSE FEATURES
a.	ROOT TERM	3
b.	"NBR TERM (FN)"	1,3
	TERM#	2
c.	EXPAND (FN=) TERM	1,2,3,4,5
d.	EXPAND TERM (,FIELD=FN)	1,2,3,4,5
e.	TERM*	1,3
f.	BROWSE FN TERM	1
g.	TERM***	1,5

4.4.8 The user, after having received a display of alphabetically adjacent terms, can have additional prior or subsequent terms displayed.

a. ORBIT (NBR)
b. ORBIT (#)
c. DIALOG
c. RECON
c. NASIS
c. DATA CENTRAL
d. SPIRES
e. STAIRS

	PROMPT	FORWARD	BACKWARD	STOP
a.	UP N OR DOWN N?	DOWN N	UP N	NONE
b.	CONTINUE PRINTING?	YES		NO
c.		PAGE +	PAGE -	
d.	MORE?	YES,or F	B	NO
e.		+KEYWORD	-KEYWORD	

4.4.9 The user can ask for a display of terms semantically related (broader, narrower, synonyms, see also, statistically associated) to the one he designates. Possible features of a display are 1) that codes are used to designate the type of relation holding between the display term and the one entered, 2) that the terms are given identifiers so that they can be referred to later, 3) that each term indicates how many records would be retrieved if it were entered in a search request, and 4) that each term indicates how many terms are cross-referenced with it.

a. ORBIT
b. DIALOG
b. RECON
c. SPIRES
d. LEADER

	COMMAND	RESPONSE FEATURES
a.	"TREE TERM"	both one level above and two levels below
b.	EXPAND E# or R#	1,2,3,4
c.	THESARUS index TERM	all terms one level below this one
	SYNONYM index TERM	all synonyms at this level
d.	ASSOCIATION or ASSOCIATED PHRASES	2,3 for phrases statistically associated with those selected so far

4.4.10 The user, when incorporating either alphabetically or semantically related terms into his query, 1) can ask for ALL of them, 2) can string them together with commas, 3) can use a hyphen to specify a range of them, and 4) can incorporate them directly into a boolean expression.

a. 2,3,4
b. 2,3
c. 2,3
d. 1,2
e. 1,4

a. NASIS
b. RECON
c. DIALOG
d. LEADER
e. ORBIT

4.4.11 The user is automatically told when he enters a search term which other terms are related to it.

BASIS

4.4.12 The user can specify that all terms hierarchically below the designated one are to be incorporated into his response.

a. EXPLODE TERM
b. available in some data bases by automatic index enhancements

a. ORBIT
b. DIALOG

4.4.13 The user can specify, as part of his request, that certain terms are to be treated as synonymous.

a. TERM1 (TERM2)
b. TERM1 SYN TERM2

a. LEADER
b. STAIRS

The following seven items facilitate limiting the scope of concepts.

4.4.14 The user can specify the case characteristics (all capital letters, little letters) that terms in the records must satisfy. — STAIRS

term .X. where X can be L/C for lower case, U/C for upper case, F/C for first letter capitalized, and M/C for mixed upper and lower case.

4.4.15 The user can specify how central the terms in a query must be to the content of a record. — a. INTREX; b. DIALOG

a. RANGE X where X can be MAJOR, SECONDARY, MINOR, or TOOL.
b. LIMIT X1/X2 where X1 is the query set or sets being qualified and X2 is either MAJ or MIN.
SELECT TERM/FN*

Often data bases star central terms (tungsten*) thus alleviating the need for a special feature at query time, but requiring the user's item-by-item post-selection of results.

4.4.16 The user can specify that words are to appear one before the other in the text of a record. An asterisk implies sequential searching. — a. ORBIT; b. DIALOG; c. RECON; d. STAIRS; e. DATA CENTRAL; f. RIQS; g. INTREX

*a. STRINGSEARCH (FN) : TERM1 TERM2 :
b. TERM1 (W) TERM2
c. 'TERM1 TERM2'
d. TERM1 ADJ TERM2
e. TERM1 TERM2
*f. 'TERM1 TERM2'
g. TERM1 - TERM2
*g. RESTRICT FIELD NAME TERM1 TERM2

4.4.17 The user can specify that words are to appear within n words of each other in the text of a record. A - means that the second word must appear before the first, and a + means that the second word must appear after the first.

a. DATA CENTRAL
b. DIALOG
c. RECON

	+	-	+ -
a.	TERM1 (WPn) TERM2	(WMn)	(Wn)
b.	TERM1 (nW) TERM2		
c.	TERM1+*n TERM2	-*n	*n

4.4.18 The user can specify that words are to appear within some logical grouping of a record.

a. DIALOG
b. STAIRS
c. INTREX
d. ORBIT

a. TERM1(F)TERM2 in same field.
   TERM1(nS)TERM2 within n sentences of each other or (if n is omitted) in the same sentence.
   TERM1(L)TERM2 in same index linkage
b. TERM1 WITH TERM2 in same sentence
   TERM1 SAME TERM2 in same paragraph
c. WITH TERM2 in same index phrase
d. SENSEARCH (FN) :TERM1 TERM2: in same sentence

4.4.19 The user can specify that the terms are to appear within a particular field (or index). FN means that a field name contains only two characters.

a. INTREX
b. RIQS
c. NASIS
d. SPIRES
e. BASIS
f. DATA CENTRAL
g. STAIRS
h. RECON
i. DIALOG
j. ORBIT

a. SUBJECT TERM
   AUTHOR TERM
   TITLE TERM
b. IF #FIELD1, #FIELD2, #FIELD3 CONTAIN. . .
c. search request, FIELD = FIELD NAME
   FIELD NAME relational operator TERM
d. FIELD NAME TERM
   FIELD NAME relational operator TERM
e. FN TERM or FIELD NAME, TERM
f. $FIELD NAME relational operator TERM
g. TERM.FIELD1, FIELD2, FIELD3.
h. FIELD NAME = TERM
i. FN = TERM indexed fields
   TERM/FN1,FN2 full text fields
j. TERM(FN)

4.4.20 The user can specify that a particular record is to be retrieved, without its being immediately displayed.

a. $NR EQU RECORD#, RECORD#. . .
b. KEEP RECORD#
c. RECORD#
d. DOCUMENT RECORD#
e. BEGIN SEARCH OF RECORD RECORD#
f. ..SELECT ALL RECORD EQ #

a. DATA CENTRAL
b. DIALOG
c. ORBIT
d. INTREX
e. RIQS
f. STAIRS

4.5 INTERCONNECTING CONCEPTS

Where large data bases are involved, a query consisting of a single concept will generally retrieve too many records. On the other hand, a query requiring the presence of four or five concepts is likely to retrieve too little. Features are needed that help the user combine concepts incrementally until as many restrictions have been imposed as possible, without impairing the volume of retrieved records. The boolean operator AND linking two terms narrows the scope of the search to only those records satisfying both conditions. The boolean operator OR linking two terms broadens the scope of the search to all those records satisfying either condition. The OR is often used to link synonyms together, while the AND is used for intersecting distinct concepts. While the features discussed in the previous section act like ORs and ANDs, they generally are processed before boolean operators. The one exception is DATA CENTRAL's word proximity operator, where FOOTBALL OR BASEBALL (W2) UMPIRE means that either football or baseball must occur within two words of umpire.

Many of the problems that might arise when mixing boolean operators are alleviated by the use of request sets. The user is able to incorporate sets into other sets, thus nesting the parts of the overall query. When he finds that a particular set needs revision, he can easily replace it with a new set. One advantage of breaking a potentially complex request into simple components is that one receives retrieval feedback after every step. When something goes wrong, like misspelling a word, the source of the difficulty is immediately apparent and can be easily corrected. LEADER is the only one of the systems that has neither request sets nor boolean operators. While the LEADER user cannot precisely specify concept interconnections, his

results are ranked beginning with those matching on the greatest number of phrases. He can cut off the display of records as the percentage meeting his relevancy criteria gets too low. Unfortunately, there is no method for breaking a request into parts that can be refined separately and later integrated.

Two areas where systems might be improved involve search profiles and sequential searching. Search profiles are often thought of as standing requests for incoming records. It is probably more economical to carry out selective dissemination of information (SDI) via batch rather than interactive searching. However, if profiles are thought of as highly polished request components for use by people searching in inadequately indexed areas, then it makes sense to have them available interactively. In many ways this use parallels publication of scholarly research and demands the same type of peer review. Not only would the task of searching be made easier, but users would effectively be indexing material in their own discipline. Sequential searching also has typically been thought of in terms of batch systems. However, when it is limited to a small set of results, it can be carried out rapidly and for a reasonable cost. Many fields in a record are used infrequently and do not warrant the cost of indexing. Yet without a sequential searching capability, they cannot be used during searching at all.

The next four items deal with boolean operators.

4.5.1 The user can interconnect portions of his request with boolean operators (connectors).

	AND	OR	AND NOT	OR NOT
a.	AND,&	OR	AND NEQ AND /=	OR NEQ OR /=
b.	*,&	+,|	-,¬	
c.	AND	OR	AND NOT	
d.	*,&,AND	+,| ,OR	-,¬,AND NOT,NOT	
e.	AND	OR,XOR	NOT	
f.	AND,&	OR,|	AND NOT,&¬	
g.	AND	OR	AND NOT	OR NOT
h.	AND,.AND.	OR,.OR.	AND NOT .AND. .NOT.	
i.	*,AND	+,OR	-,NOT	

a. DATA CENTRAL
b. NASIS
c. ORBIT
c. INTREX
d. RECON
e. STAIRS
f. SPIRES
g. RIQS
h. BASIS
i. DIALOG

4.5.2 For systems permitting mixed expressions, connectors with high priority are processed before connectors with low priority, and connectors with equal priority are processed from left to right. The features discussed in the previous section either cannot be mixed in with boolean operators or else have higher priority. The one exception is DATA CENTRAL's word proximity operator which has lower priority than OR so has been included below.

a. DATA CENTRAL
b. NASIS
c. ORBIT
d. DIALOG
d. RECON
e. STAIRS
f. SPIRES
g. RIQS
h. BASIS
i. INTREX

	HIGH PRIORITY		LOW PRIORITY
a.	&	OR	W,WP,WM,AND
b.	*,&,-,¬	+,|	
c.	AND,AND NOT	OR	
d.	*,-,AND,NOT	+,OR	
e.	AND,NOT	OR,XOR	
f.	AND,&,OR,|,AND NOT,&¬		
g.	AND,AND NOT	OR,OR NOT	
h.	NOT .NOT.	OR .AND.	AND .OR.
i.	AND,OR,AND NOT		

4.5.3 The priority of connectors can be controlled through the use of parentheses.

NASIS
DIALOG
RECON
STAIRS
SPIRES
RIQS
BASIS

4.5.4 A boolean operator can be specified for insertion between terms in a range of terms.

a. set# - set#/operator
b. TERM1-TERM2/operator
   operator/TERM1:TERM2
c. TERM1 operator TERM2 TERM3

a. DIALOG
a. RECON
b. NASIS
c. STAIRS

The next eleven items deal with request sets.

4.5.5 Query input from the user is treated as a complete request. It is possible 1) to store the results of a request and 2) to invoke those results as the domain (or part of the domain) for a later query.

a. 1) IF query THEN PLACE RECORD IN SET #
   2) BEGIN SEARCH OF SET #
      BEGIN SEARCH OF UNION (#,#,...,#)
      BEGIN SEARCH OF INTERSECTION (#,#,...,#)
      BEGIN SEARCH OF COMPLEMENT (#,#,...,#)

a. RIQS

4.5.6 Query input from the user either begins a new request or else continues the current one. While it is possible to back up and take another tack (for INTREX backup points must have been named previously), it is not possible to compose new requests by connecting together previous search results (INTREX is an exception).

a. SPIRES
b. DATA CENTRAL
c. LEADER
d. INTREX

	FIRST PART	CONTINUATION	BACKTRACK
a.	FIND query	AND or OR query	BACKUP (one step)
b.	query	AND or OR query	MODIFY # (to step #)
c.	query leads to phrase selection	PHRASES for phrase selection revision	START OVER for query revision while maintaining selected phrases NEWSEARCH for total query revision
d.	FIELD NAME query	AND or OR query	SET2 (to bring back SET2)

4.5.7 Query input from the user is treated as a complete request and is assigned a set number (except for INTREX). The results of a request can be incorporated into subsequent requests by mentioning the set number (or set name for INTREX).

ORBIT
DIALOG
RECON
STAIRS
BASIS
NASIS
INTREX*

* result sets, or backup points, are established via the command NAME SET1 where SET1 is to be the name of the set.

4.5.8 The only queries that can contain connectors are those where the only terms are set numbers.

a. DIALOG
b. BASIS

a. COMBINE query
b. (query)

4.5.9 The user can review the text of his query (or queries).

a. ORBIT
b. DIALOG
c. RECON
d. STAIRS
e. DATA CENTRAL
f. BASIS
g. LEADER
h. NASIS

a. "DIAGRAM#" The text of the query for set number # (or for the most current set if # is omitted) is displayed, but with all set numbers replaced by query text (enclosed in parentheses).
b. DISPLAY SET HISTORY For each active set the text of the query (with index phrases replacing EXPAND display line numbers) is listed along with the number of records the query retrieved.
c. HISTORY Same effect as for DIALOG.

d. ..DISPLAY ALL or ..DISPLAY #1,#2 For each set in the range specified, the text of the query is listed along with the number of records the query retrieved.
e. REQUEST or REQUEST #, ANSWER # or ANSWER. For the latest addition to the query or for the designated modification step, either the query text or else the number of retrieved records is displayed.
f. /LIST ALL. Same effect as for STAIRS.
g. PHRASES The index phrases selected so far are listed along with the number of records each retrieves.
h. SETS. Same effect as for DIALOG.

4.5.10 The final set is used for accumulating individual records as they are being displayed.

a. DIALOG
b. RECON

a. KEEP record# or
KEEP set#/item#, item#-item#, ...

4.5.11 Before submitting any queries, the entire data base acts like result set 0, and can be displayed.

STAIRS
NASIS

4.5.12 Result sets can be purged so as to free space for additional sets.

a. STAIRS
b. RIQS
b. RECON
c. ORBIT
d. INTREX

a. PURGE ALL or PURGE #, or PURGE #,#
b. RELEASE SET # or RELEASE SETS #,#,...,#
c. ERASEALL or ERASEBACK # (# and subsequent sets are wiped out) or RESTACK n,...,m TO # (# and subsequent sets are wiped out except for n,...m which are given new set #s beginning at #. If the TO # part is omitted set 1 is assumed.
d. DELETE SET# SET#

4.5.13 Result sets can be saved for re-use during a later session.

a. RIQS
b. INTREX

a. if, during logoff, the logical file used for storing was EXTENDED, then the sets are available as part of the data base.
b. if have OPEN FILE FILE1 then SAVE SET1 SET2 will have sets with those names across sessions and USE FILE1 will restore them.

4.5.14 The total request can be saved for re-use during some other session.

a. STAIRS
b. NASIS
c. SPIRES
d. DIALOG

	SAVE	USE	DELETE
a.	..SAVE NAME	..EXEC NAME	..PURGE NAME
b.	PROFILE	RERUN NAME	
c.		EXECUTE strategy in working memory	
d.	END/SAVE a serial # is assigned	.RECALL #	.RELEASE #

4.5.15 Certain commands (usually including the AND NOT connector) can be used only in conjunction with a request set (or ongoing request).

a. STAIRS
b. INTREX
c. ORBIT

a. ..SELECT the system puts out a set number and the user responds with a prior set number (or ALL) followed by query input. Numeric and other non-indexed fields can only be searched in this manner.
b. RESTRICT FIELD NAME TEXT.
c. STRINGSEARCH set# query

5.0 RESULT MANIPULATION FEATURES

5.1 INTRODUCTION

The fact that records in a data base can easily be manipulated by a computer means that records can take on different appearances depending upon the needs of the user. During query formulation, aspects of records can be presented that help the user revise his request and measure his progress. After a query has been formulated, other aspects can be presented that help the user locate answers to his questions. Occasionally, aggregate properties of records are more important than individual records. In this chapter we shall discuss features for interactively displaying records during and after the query negotiation process.

The result manipulation features available in the eleven systems are not as extensive as those needed in management information systems. Ideally, data bases intended for decision making should reflect problem-oriented elements and relationships. It should be possible for the user to organize elementary components so that they reveal strategies for dealing with the problem. Most data bases of the eleven systems reflect a record orientation rather than a problem orientation. Records entering the data base remain identifiable, and generally are the same records that are retrieved later on. While NASIS, DATA CENTRAL, and SPIRES can synthesize retrieved records out of stored record constituents, they are not commonly used in this way.

For data bases to be developed that effectively address problems, both searchers and data base contributors must agree upon problem conceptualizations. There must be some guarantee regarding the quality of queries and retrieved records. While for some RIQS and SPIRES data bases the contributor

and searcher are a single individual, and for some NASIS, DATA CENTRAL and BASIS data bases the contributors and searchers are employees of the data base owner, these are exceptional situations. Generally, the only thing different records have in common is the record format. Searchers rarely have any contact with data base suppliers. It is by making comprehensive data bases available to very large user populations that interactive searching proves economical.

5.2 DISPLAYING RESULTS SIMPLY

Different systems make different assumptions about how much a user should specify for a display of retrieved results, and what should be done in case he does not specify. Asking for a display is much simpler than formulating a query, and the inexperienced user probably does not feel any need for control over the process.

5.2.1 The system informs the user that search results can be displayed, and tells him how to ask for them. The user can ask for a standard display at other times by typing OUTPUT, PRINT, or TITLES respectively.

a. INTREX
b. DATA CENTRAL
c. LEADER

a. "A search on your request SUBJECT so-and-so found # documents. The catalog fields TITLE, AUTHOR, LOCATION for those documents will be output when you type o (for output)...."
b. # ENTRIES,
DO YOU WANT TO PROCESS ANSWERS
NO, PRINT OR MODIFY?
c. After the user STOPs phrase selection the system responds
YOU HAVE SELECTED # PHRASES. WOULD YOU LIKE ME TO START RETRIEVING TITLES, NOW.

5.2.2 The user is assumed to know how to ask for a standard display of the most recent result set.

a. DISPLAY set#
b. TYPE (or OUTPUT if listing off-line later)
c. "PRINT"
d. ..BROWSE
e. DISPLAY (for the CRT) or TYPE (for typewriters)
f. IF query THEN DISPLAY fields; END (or PRINT, not DISPLAY, if listing off-line later)
* the results of a request are not displayed immediately. The user is asked DO YOU WANT THE DISPLAY REPORTS LISTED

a. BASIS
a. NASIS
b. SPIRES
c. ORBIT
d. STAIRS
e. DIALOG
e. RECON
f. RIQS*

5.2.3 The user will receive a single record or the first page of a multipage record and then will be asked to indicate if he wishes to continue.

a. TYPE-GO, NEXT (FOR NEXT DOC), X (FOR DOC X), STOP OR TERMS
b. PAGE
* for CRTs only

a. LEADER
b. RECON (DISPLAY)
b. DIALOG
b. NASIS
b. DATA CENTRAL*

5.2.4 The user will receive a display of up to five records and then will be asked if he wishes to continue.

a. CONTINUE PRINTING? (YES/NO)
b. The user is warned that the output will be lengthy and may select how many documents he wants to see before being interrupted.

a. ORBIT
b. BASIS

5.2.5 The user will receive a continuous display of records that can be interrupted (or is interrupted automatically at page boundaries by the terminal controller).

* for hard copy terminals only

SPIRES
INTREX
RIQS
RECON (TYPE)
STAIRS
DATA CENTRAL*

5.2.6 The user is displayed fields from each record in accordance with 1) a system level predefined format, or 2) a data base specific predefined format. The format is likely to contain 3) all the fields in a record that the user is permitted to see, 4) a logical subset of the fields.

a. 2,4
b. 1,4
c. 1,3

a. ORBIT
a. DIALOG
a. DATA CENTRAL
b. LEADER
b. INTREX
b. NASIS
c. RECON
c. SPIRES

5.2.7 The user must specify whether all the fields or some of the fields are to be displayed. The STAIRS user is provided with a list of the possible fields only if he asks for an explanation. The BASIS user is provided with a list immediately before the display begins. The RIQS user is provided with a list of the possible fields only if he asks for one during the system start up phase.

a. *PLEASE CHOOSE PARAGRAPHS: ENTER A-ALL,
B-PARAGRAPH LIST, OR C-EXPLAIN
b. THE DATA ELEMENTS FOR THE ___ DATA BASE ARE
#-FIELD NAME
WHAT FIELDS DO YOU WANT TO SEE?
ENTER FIELD NUMBERS SEPARATED BY COMMAS OR ALL
c. The fields can be like #4, #8 THRU #12,
#AUTHOR for a subset, or RECORD for all.

a. SPIRES
b. BASIS
c. RIQS

5.2.8 The records are displayed a) with the earliest entered appearing first, b) with the latest entered appearing first, c) with the order different for different data bases, or d) with those most closely matching the query appearing first.

a. STAIRS
a. NASIS
a. BASIS
a. RIQS
b. ORBIT
b. INTREX
b. DIALOG
c. SPIRES
c. DATA CENTRAL
c. RECON
d. LEADER

5.2.9 The records are displayed with the terms causing retrieval highlighted in some manner (also see 5.3.9).

a. RECON
b. LEADER
c. DATA CENTRAL
c. STAIRS*

a. words in the body of the text causing retrieval are enclosed in asterisks.
b. the index phrases causing retrieval are starred.
c. users must specify before searching whether or not there should be highlighting.
* only on display terminals (3270's).

5.3 CONTROLLING THE DISPLAY

The more experienced a user, the more likely he is to feel a need to control the display. If he is looking for additional terms to include in the request, he may only want to see fields containing key words. He may want to look at abstracts after discovering that the title and index terms look promising. When he finds a record that he feels should not have been retrieved, he may want to get rid of it and/or find out why it was retrieved. His primary concern is that the display be rapid, pertinent, and easily interrupted. When the user begins to treat his results as a finished product, his interests may change. Brevity becomes less important than completeness. While system representatives indicate that most users treat printed results more like grocery lists than like manuscripts, some users do worry about page layout and sequencing. Given the variety of display needs, some control must be placed in the hands of the user, but it is not clear how to provide flexibility simply. DATA CENTRAL and STAIRS provide the report generation capability needed for photocomposition, but only to the off-line user. SPIRES provides the on-line user with the full power of a text-editor, and INTREX has a special terminal with hardware and software editing features, but these are probably more than is needed. Perhaps the best approach is a mixture of predefined formats and features for on-line construction of simple formats.

5.3.1 The user can have the records displayed on-line using a predefined format that differs from the default format or from those discussed below.

a. DATA CENTRAL
b. NASIS
c. DIALOG
c. RECON

a. if user response to NO, PRINT, MODIFY is OUTPUT, then system prompts ENTER DESIRED OUTPUT, DEVICE where DESIRED OUTPUT can be the name of one of the data base's predefined formats.
b. DISPLAY set# , format#
c. DISPLAY set# /format#

5.3.2 The user can specify that every field (that he has access to) of the records is to be displayed.

a. DATA CENTRAL
b. NASIS
c. INTREX
d. DIALOG
e. RECON
f. STAIRS
g. RIQS
h. BASIS
i. ORBIT

a. after the prompt ENTER DESIRED OUTPUT, DEVICE the response FUL-RET, CONSOLE
b. DISPLAY set#, 4
c. OUTPUT ALL
d. DISPLAY set# /5
e. DISPLAY set# /2 (this is usually the default)
f. ..BROWSE ALL
g. IF query THEN DISPLAY RECORD; END
h. DISPLAY set# causes the prompt WHAT FIELDS DO YOU WANT TO SEE? The response is ALL
i. "PRINT FULL"

5.3.3 The user can specify that only certain fields are to be displayed. An asterisk means that the order in which the fields are specified by the user determines the order in displayed records.

a. DATA CENTRAL*
b. INTREX
c. STAIRS
d. SPIRES
e. RIQS*
e. BASIS*
f. ORBIT*
g. NASIS*

a. after the prompt ENTER DESIRED OUTPUT, DEVICE users says ASK-ME, CONSOLE system responds ENTER SEGMENTS TO BE DISPLAYED IN ANSWERS user responds FIELD NAME, FIELD NAME,...
b. OUTPUT FIELD NAME, FIELD NAME,...
c. ..BROWSE FIELD NAME, FIELD NAME,...
d. SET ELEMENTS FIELD NAME, FIELD NAME,... limits only that command.
e. see 5.2.7
f. "PRINT INCLUDE FN, FN EXCLUDE FN, FN" to modify the standard set of displayed fields. Neither the INCLUDE portion nor the EXCLUDE portion is required.
g. FORMAT #, $FIELD NAME, $FIELD NAME

5.3.4 The user can indicate how records are to be displayed by including parameter values within the display command.

a. ORBIT
b. RIQS
c. INTREX

a. "PRINT INDENT" to receive full names of fields.
b. DISPLAY ACROSS to have more than 1 field/line
SUPPRESS to delete field names
TAB # to skip to a column
INDENT # to tab when continuing on a new line.
SKIP # to leave blank lines
WIDTH # to specify output width
SPACE # to leave blank spaces
/ to skip to a new line
EJECT to skip to a new page
c. OUTPUT FAST displays just the first line of output for each record.

5.3.5 The user can develop formats on-line for later incorporation into record displays.

a. RECON
b. STAIRS
c. NASIS
d. BASIS

a. FORMAT FN, FN;FN (commas indicate line continuation, semicolons indicate carriage return).
b. ..PAGE LINES for text length per page
MARGIN upper margin size
LENGTH for text width
START for skipping to the next page.
c. FORMAT F set # (S,P) FN (#,T)
(S or (P for screen or printer width
,P)or,NP) for page or no page number
(# to skip to a column
,T) or,S) or ,A) to tally, sum or average column values.
HEADER FN,FN to put headings on columns
d. SET LINE LENGTH
HEADINGS (ON/OFF)
FIELD NUMBERS (ON/OFF)

5.3.6 The user can designate the order in which records are to be sorted.

a. DATA CENTRAL
b. RECON
c. DIALOG
d. STAIRS

a. after the user specifies output fields, the system asks
DO YOU WANT THE ENTRIES SEQUENCED, YES OR NO?
if YES then ENTER SORT SEGMENTS (LENGTH, MODE)
FIELD NAME, # characters, A or D
b. SORT set# / FN, A or D, field length / etc.
c. PRINT set/format/item range /FN, A or D/etc.
. SORT set/item range /FN, A or D
d. ..SORT FIELD NAME A or D (200 records or less)

5.3.7 The user can designate that the records are to be ranked.

a. ORBIT
b. STAIRS

a. PRINT HITS for displaying documents ranked according to degree of query-index term matching.
b. ..RANK set# (200 records or less)

5.3.8 The user can calculate sums, averages, or other values for subsets of records.

a. BASIS
b. SPIRES
c. RIQS
d. NASIS

a. DISPLAY for some data bases asks ENTER COMP EXPRESSION.
Values to be displayed can be products of fields and/or constants, and can be accumulated. For example
3,12, MILES*".13"/GALLONS, CUM (MILES)
prints four values per record.
b. AVERAGE FIELD NAME calculates two averages--one for the field's values in the entire (small) data base and one for the field's values in the retrieved subset.
c. LET variable = arithmetic expression incorporating fields and/or constants. This can occur before, during, or after a search.
d. see 5.3.5

5.3.9 The user can specify that he wants to have the display focus upon the term(s) in the record that caused the record to be retrieved.

a. DATA CENTRAL
b. INTREX
c. RECON
d. STAIRS

a. When selecting the data base, KWIC must be specified as one of the options if words causing retrieval are to be highlighted during display.
After record output begins, $\pm$ # KEYWORDS causes a predefined number of adjacent words to be included in the display.
b. OUTPUT MATCH FIELD1 FIELD2 limits the display to index phrases causing retrieval along with the specified fields (FIELD1 FIELD2).
c. After record output begins, PAGE ++ or PAGE -- causes skipping forward or backward to the next page containing a term that led to retrieval (the term is set off by asterisks).
d. A display highlight switch can be set by the user.

The following three items deal with features for changing display directions in midstream.

5.3.10 The user can have a display begin at the nth record in the subset. In this way, he can discontinue display, switch to a new format, and resume where he left off.

a. RECON
a. DIALOG
b. NASIS
c. STAIRS
d. LEADER
e. ORBIT
f. BASIS
g. DATA CENTRAL

a. DISPLAY set#/format#/item# or item#-item#
b. DISPLAY set#,format#,item#
c. After record output begins,
   DOC = item#, item#
d. After title page output,
   item#
   causes display to skip to that record position
e. "PRINT SKIP item#"
f. After selecting fields to be displayed, the system asks which records the user desires to see. Possible skipping responses are
   item# to item#
   # of items, item#
g. After record output begins,
   +# ENTRIES

5.3.11 Without going through the process of selecting a new format, the user can specify that he would like to see additional portions of the record he is currently viewing.

a. LEADER
b. RECON
b. NASIS
c. DIALOG
d. BASIS
e. DATA CENTRAL
f. INTREX

a. during output TERMS causes a record's index phrase to be displayed.
b. DISPLAY accession#
c. DISPLAY accession#/5
d. DISPLAY (accession#) followed by a selection of ALL field names.
e. +# KEYWORDS to broaden the window around a highlighted term.
f. DOC accession#/OUTPUT FIELD1...

5.3.12 During the display, the user can easily specify which records he would like to have kept.

a. SPIRES
b. RECON
c. DIALOG
d. INTREX

a. the TYPE KEEP command displays records one at a time and asks
KEEP?
after each one. Records the user states he does not want are purged from his request set.
b. KEEP record#
c. KEEP keeps the record currently being viewed
KEEP set#/item# or item#-item#
d. DOC accession#/OR SAVESET

The following three items deal with features for printing.

5.3.13 The user can specify that the retrieved records are to be printed offline.

a. DATA CENTRAL
b. NASIS
b. DIALOG
b. RECON
b. BASIS
c. INTREX
d. ORBIT
e. STAIRS
f. SPIRES
g. RIQS
h. LEADER

a. when responding to
ENTER DESIRED OUTPUT, DEVICE.
the device PRINTER can be specified.
b. PRINT with the same options as the DISPLAY command
c. OUTPUT OFFLINE in place of OUTPUT
d. "PRINT..." with OFF-LINE as the final option.
e. ..MAIL request set(s) field name list
where ...MAIL without options means print the most recent set using the same field name list as specified in the latest ..BROWSE command. Or
e. ..PRINT record# range where PRINT without options means the document most recently displayed. record numbers are relative to the beginning of the most recent request set.
f. LIST OFFLINE BIN = #
causes those records displayed via OUTPUT to be printed and routed to the appropriate location.
g. DIVERT, PRINT or DISPLAY, printing site
causes those records displayed via PRINT or DISPLAY to be printed at the appropriate location.
h. after STOPping the display of records, the user is asked
DO YOU WANT TO PRINT ANY TITLES.
If the user responds YES, he is asked
INCLUDING THE ABSTRACTS.
Then he is asked to specify which records in the set are to be printed.

5.3.14 The user can specify where printed results are to be mailed.

a. BASIS
b. ORBIT
c. RECON
c. DIALOG
d. NASIS
e. INTREX

a. after requesting printing, the user is asked
PLEASE TYPE COMPLETE NAME AND
MAILING ADDRESS
b. after requesting off-line printing, the user is asked
NAME AND BUSINESS PHONE?
ADDRESS?
CITY, STATE (PLEASE DO NOT ABBREVIATE), AND ZIP -
REQ NAME, OR SAME
TITLE, OR NONE
and then is asked if the above is OK?
c. before ever beginning to search, the user is asked
SEARCH TITLE
NAME OF PERSON CONDUCTING SEARCH
NAME OF PERSON RECEIVING RESULTS
MAIL ADDRESS
d. before ever beginning to search
ENTER YOUR NAME:
ENTER YOUR ADDRESS:
e. the user so specifies in a NOTE or COMMENT

5.3.15 The user can specify what heading should appear on his printed output.

a. NASIS
b. INTREX
c. LEADER
d. RIQS
e. ORBIT
e. DIALOG
e. RECON

a. TITLE followed by the heading
b. NOTE followed by the heading
c. after indicating which documents to have printed, the user is asked
WOULD YOU LIKE TO LABEL YOUR OUTPUT.
If YES, then the user is repeatedly asked
ENTER LINE OR STOP -
d. before entering the query, the user can specify
BEFORE SEARCH DISPLAY 'heading'
e. see 5.3.14

The following three items are clusters of display features that have been incorporated into some interactive retrieval systems.

5.3.16	The user can specify that microform copies of the retrieved documents are to be displayed at the terminal.	INTREX BASIS DIALOG
5.3.17	The user can specify that a graph is to be generated from the retrieved records and is to be plotted at the terminal.	BASIS RIQS
5.3.18	The user can specify that the records are to be passed through statistical analysis, and that the results of analysis are to be displayed at the terminal.	RIQS DIALOG BASIS

6.0 CONCLUSIONS

6.1 INTRODUCTION

Most of the report has been descriptive rather than normative. It has been left to the reader to choose between alternate implementations of equivalent features. A more normative approach will be taken in this chapter. At the April 1973 workshop, designers were asked to specify which features were in some sense "minimal." What this came to mean was that, without a feature, searching would be significantly constrained. The minimal features will be summarized in the next section along with supportive reasoning.

The introduction of supportive reasoning brings up another issue. While students of design can profit from hindsight, they also need principles that lead them into good design. The new capabilities of a rapidly evolving computer technology too often are the only guideposts available. A possible set of user-oriented principles will be presented in the third section along with supportive examples drawn from the report. The principles will not be backed up with citations to a well-developed body of man-computer problem-solving literature because that body of evidence has not yet been created. Hopefully some readers will be stimulated to see if in fact the proposed principles can be confirmed or disproven.

Finally, an attempt will be made to forecast the role that interactive retrieval systems might play in the future. From the forecast, recommendations will be derived regarding the advantages of various forms of system coupling and/or standardization. It is still too early in terms of technological evolution to determine how best to build a system, but it is

not yet too early to begin deciding why we want to build an information infrastructure and for whom it is intended.

6.2 MINIMAL FEATURES

Twelve of the features listed in the appendix were found by designers at the April 1973 workshop to be of central importance to interactive searching. Two of the features relate to instruction and assistance, seven relate to formulating queries, and three relate to manipulating the retrieved records.

The two features felt essential for instruction and assistance are live help and complete, readable users' guides. That designers consider human and written assistance to be essential is a clear indication that at present these traditional media are more necessary to searching than is assistance via the computerized medium. One should not conclude, however, that on-line documentation and assistance are unimportant. Manuals and live help derive much of their value from the fact that they can be turned to when all else fails. Computers have been used for instruction and assistance for only a very short time. Authors, instructors, consultants, and those in need of help are far more comfortable with traditional media than they are with the new ones. It is not surprising that mistakes are made by both preparers and users of on-line materials, or that people rely upon the traditional channels even though the computerized ones could be used. As better materials are developed for supporting users on-line, and as more people acquire computer literacy, then print and human contact will become less essential for effective searching.

The minimal features for query formulation are far harder to categorize. Of the seven, only four are strictly necessary for constructing a request. These are boolean operators, search field control,

relational operators, and suffix removal. Of all the features, boolean operators is probably the most essential feature. Without ANDs, ORs, and NOTs it would be difficult to formulate complex requests. Of the eleven systems only LEADER does not provide for them. The indexing phrases selected by a LEADER user are assumed to be additive; retrieved documents are ranked in accordance with the additive overlap between request and document. Ranking accomplishes most of what normal users expect from a system, but definitely limits what can be expressed in a query. There is no way to indicate to LEADER that documents containing extraneous concepts should be demoted in the ranking, or that documents containing central concepts should be promoted. There is no way to indicate that two or more concepts are not independent. For example, different phrases may stand for a single concept, but this cannot be specified. The net effect is that without boolean operators the skilled searcher cannot retrieve exactly what he wants.

Control over which fields to examine when matching requests with stored records is also essential. Like people, records tend to have multiple facets and rarely can be treated in just one way. By establishing fields, and by making sure that each field contains only a single type of information, it becomes possible to compare records along unitary dimensions. In order for the searcher to take advantage of the various fields, he must be able to refer to them in the request. He can not only specify individual fields he wants to interrogate, but he can usually collapse many fields into a single field. All of the systems except LEADER allow the searcher to specify the fields he wants to interrogate. The LEADER user is limited to searching via index phrases. While this does not impede the beginning user, it frustrates the person who knows about

fields and wants to take advantage of them. SPIRES and INTREX do not allow the user to combine fields. This means that the user is forced to include field names in his requests even though he sees no reason for doing so.

Relational operators and suffix removal are really a single feature. They make it possible to treat a range of values as if it is a single value. While one could OR together individual values within the range, it would be far too tedious to do so. Suffix removal is used primarily for retrieving ranges of words that contain a single stem. Relational operators are used primarily for retrieving ranges of numeric data. Unlike words sharing a common root, the limits of a numeric range rarely share common digits and so both values must be specified. Of the eleven systems, only BASIS has neither suffix removal nor relational operators. NASIS and RIQS deal with textual information but do not have suffix removal.

The three remaining query formulation features are not needed in a strictly logical sense but do contribute to better searching. They are dictionary access, request set, and search review. Providing access to the alphabetically arranged list of words and/or phrases that retrieve records responds to a number of behavioral needs. Many users are not good spellers and are not likely to retrieve what they want unless they have access to a dictionary. While it is difficult to find words in a dictionary that one does not know how to spell, most bad spellers at some point have mastered this art. Most systems provide frequency counts along with the words, and these help users avoid indexer and author spelling errors. Frequency counts also help the skilled searcher select terms. Generally searching is most effective when the request does not depend upon heavily or rarely used words. The major reason for providing access to a dictionary of indexing phrases is that it allows the user to shift from recall to recognition. It is well known to psychologists that

people recognize far more than they can recall. Many searchers forget to OR together slight variants of the same word. The dictionary display reminds the searcher that there are many words beginning with the same stem but having different suffixes. Of the eleven systems, only LEADER, RIQS, and INTREX do not have dictionary display. While the LEADER user is deprived of spelling assistance, the extensive display of phrases containing query words and the display of phrases statistically associated with selected phrases more than compensate for this lack.

Request sets and search reviews also respond to a number of behavioral needs. By changing query formulation from a single process into a set of processes which can be tied together, request sets both simplify the task and lessen the need for preplanning. The searcher knows that he can switch to browsing through documents and can pursue interesting tangents suggested by the documents. Query fragments that do not prove fruitful do not need to be included in the final request. Many creative individuals rely upon apparently random thinking for discovering valuable new ideas, and it is important that they be allowed to carry familiar habits over to interactive searching. Other individuals think in a more structured manner, breaking complex problems into manageable subproblems. Request sets encourage this approach to searching as well, and free the searcher to focus upon subproblems without having to concern himself with the overall problem. Search review is necessary in conjunction with request sets, because it allows either type of individual to step back and see what he has done so far. The person at the terminal is so involved in the immediate present that he loses track of his prior actions. Unless he can remind himself of his earlier discoveries, he may forget to incorporate them into the final result. SPIRES is the only one of the systems that has neither request sets nor search review. LEADER and DATA CENTRAL

have search review though they do not have request sets. This suggests that there is a value in providing for overviews of progress to date even though the only way to change directions is to backtrack and take another approach. INTREX and RIQS have request sets but not search review. Both of these systems consider request sets to be a method for carrying work from one session at the terminal over to the next, and so do not consider them to be an integral part of query formulation.

Finally, three features are felt essential to the manipulation of search results. The user should be able to rely upon predefined formats for displaying results, should be able to construct simple display formats at the terminal in case predefined formats prove unsatisfactory, and should be able to have results printed off-line. The first two features need little elaboration. LEADER is the only system that constrains the user to predefined formats. STAIRS, BASIS, and RIQS force the user to specify which fields he wants to have displayed. The interesting feature is off-line printing. Every system provides the capability. This suggests that the information searchers retrieve cannot be easily carried away from the terminal in one's head. Having results listed at a hard-copy terminal does not alleviate the need for off-line printing. System representatives reported that users who output records at a terminal generally retrieve fewer records and display less information about each record. At present most hard-copy terminals take too long to list as much as most searchers would like to have listed. However, when time is of the essence, a little information is better than no information at all. Searchers at video terminals often bring note pads with them so that they can compensate for the missing hard-copy capability. Designers who think that electronic terminals can supplant the need for paper should think carefully about the implications of this essential off-line printing capability.

6.3 PRINCIPLES FOR USER-ORIENTED SYSTEM DESIGN

A number of principles underlying good design have surfaced in the analysis of the minimal features. To summarize, designers should: (1) rely upon traditional channels--print and personal contact--to bolster computerized communication, (2) incorporate features that facilitate the carryover of problem solving styles ranging from trial and error to rigorous analysis, (3) make it easy for the person who does not want to master the system to carry out his task, yet make it possible for the person who knows what he is doing to exercise control, (4) incorporate mechanisms that encourage users to rely upon recognition rather than recall memory, (5) avoid penalizing users for common human failings like bad spelling or lapses of memory, and (6) make it possible for users to switch back and forth from focusing upon details to capturing a sense of the entire problem.

A number of other principles are not revealed by the minimal features. For example, none of the minimal features help users discover how to search more skillfully. It would seem that designers should (7) provide mechanisms for measuring searching proficiency and (8) facilitate the sharing of user experiences suggestive of how best to search. While none of the systems have sample problems and solutions, a few have mechanisms that could be used for sharing experience. Users of DIALOG are able to construct and store queries or components of queries. A list of the stored queries is periodically circulated so that users can decide which strategies to get rid of. Those query components that many people want to use could be made publicly accessible. The net effect would be to reward those who construct good queries while allowing others to profit from their

experience. By providing for storage of and access to query components, designers make it possible for users to create searching aids for a data base tailored to the community's particular needs.

Closely related to the self-improvement principles is a principle that designers should (9) make it possible for searchers to discover why their strategies do not work so that they can learn from experience. One of the times it is easiest to make mistakes is when one is composing logically complex requests. It is difficult to determine why nothing was retrieved or why the retrieved records appear to be irrelevant unless one can see how the request was interpreted. Only INTREX and STAIRS have features for tracing search logic. With most of the other systems the clever individual can get the equivalent of search logic tracing by entering the request a piece at a time. Another way in which it is easy to go wrong is to forget that words have multiple meanings. In order to get rid of the extraneous meanings, one must first discover them. Highlighting features make it easy to zero in on the word or words causing retrieval. The user can easily discover that an irrelevant record was retrieved because of an unanticipated multiple meaning.

Many other principles might be proposed, but the few that have been discussed are sufficient to indicate what is needed. The manner in which a person interacts with a computerized system and the manner in which he carries out his searching task are not isolable phenomena. They are instances of behavior and can best be understood in a broader communication and/or problem-solving context. One can only discover what is unique about man-computer communication and interactive searching by contrasting each with its respective behavioral background. While principles can be derived from experience, there is a need to relate the principles to similar phenomena and to carry out tests that measure the reliability of the principles.

6.4 A FUTURE FOR INTERACTIVE RETRIEVAL

Interactive retrieval systems are like any other communication medium in that they have been used initially to carry the messages of the older media. They have proven their worth in the marketplace. The features needed in retrieval systems are now well understood. It is time to step back and attempt to forecast what new types of messages the medium will engender. In doing so, it is very important to remember the conclusion of the workshop participants that no distinction should be drawn between searching in bibliographic and management data bases.

One of the values of interactive searching is that many small intellectual communities both geographically and temporally dispersed can establish and maintain intra-group communication. The data base is the vehicle for simplifying access and for distributing costs over a larger community. Data bases come into existence for much the same reasons that banks do--it is easier to contribute to and withdraw from a central account than it is to maintain separate accounts with all of one's colleagues.

Interactive systems differ from traditional ones in many respects. The ability to intersect concepts and to juggle them around until the right blend has been achieved greatly expands the space within which groups can locate themselves. The ability to switch from data aggregation to single record retrieval, and to mix textual with numeric data makes it possible for searchers to see both the trees and the forest, the individual cases and the trends running through cases. The fact that terminals can be used equally well for generating new records and for displaying old ones suggests that increasingly people will use interactive systems for submitting records to data bases. There is no reason why

user communities should not be able to shape data bases to meet their needs. Retrieval systems have made it possible for searchers to get in contact with data bases, but many data bases are not really designed for comfortable searching. The burden is now upon data base managers to facilitate searching for records and to insure uniformity where it is needed for data aggregation.

We predict that in the future retrieval systems will be used heavily by small dispersed groups both to submit and to retrieve records. Currently retrieval systems are available only to the intellectually rich, but as costs go down more and more people will be able to use them for more mundane purposes. The number of groups is likely to expand and the time from submission of material to heavy usage to archiving is likely to become shorter.

We also predict that individuals will come to use a variety of different data bases, ranging from heavy use of small private files to infrequent use of large public ones. There will be a need to couple one's personal computer onto national networks. While it would be nice if one's personal searching language transferred over to the national retrieval systems, it is likely that different individuals will have quite different private languages tailored to their thinking styles, data base needs, and related computing behavior.

What is needed is not agreement regarding the syntax of retrieval languages but agreement regarding features which all systems should possess. The minimal features go a long way toward meeting this need. If the local and national systems share common functions, it should be possible to translate the syntax of one into the syntax of the other. Where automatic coupling is awkward or impossible, it is likely that

human brokers will be available who can communicate with a wide range of different types of systems.

In conclusion, if the cost of terminals and of data base storage media continue to fall, then an increasing number of people are likely to become interactive searchers and data bases are likely to proliferate. In society the impact will be to promote still more specialization. If steps are taken now to insure that systems respond to the functional needs of users, and that they share common features, then in the future it should not be difficult to couple them together into a national or international resource.

APPENDIX

At the April, 1973 workshop matrices with explanations were used to facilitate discussion between designers. The matrices summarized the degree to which each of the eleven systems incorporated various features. In October revised versions of the matrices were released as part of a comparative analysis interim report. Since that report had wide circulation and effectively summarizes the current report, the matrices and explanations have been retained as an appendix. Like the rest of the report, they reflect the state of affairs as of April, 1973. Slight changes have been made to matrix contents at the request of system representatives, but no columns have been added, reordered, or renamed. Consequently, although each matrix corresponds to a chapter, each column does not correspond to a subsection. Columns have been provided with subsection numbers so the reader can rapidly locate implementational details in the report. The matrices neither should nor can be used to rate systems. One system is not better than another because it has more check marks. Without knowledge of the specific data bases to be stored, the specific user community, and the specific operational environment, no adequate rating of systems is possible.

INTREX	RIQS	NASIS	SPIRES	LEADER	BASIS	DATAC	STAIRS	RECON	DIALOG	ORBIT	SYSTEMS SELECTED
A		b	a	A	a,b	b	B	A	A	A	LARGE TEXTUAL DATA BASES 2.3.2, 2.3.4
	X	A	X		X	A		x	x		MANAGEMENT INFORMATION 2.3.5, 2.3.6
X	x	x	x	x	x	X	X	X	X	x	FULL TEXT SEARCHING
X		X	x	X	X			X	X	X	INDEX SEARCHING
	x	X	x	x	x	x	X	X	X	X	INTERMEDIARY SEARCHER
X	X	x	X	X	X	X	x	x	x	x	END USER SEARCHER
X	x	x	x	X	x	X	X	X	X	x	VIDEO TERMINALS
x	X	X	X	x	X	x	x	X	X	X	TYPEWRITER TERMINALS
		x	x	x	X		X	X	X	X	TELEPHONE NETWORK 2.3.11, 2.3.12
X	x		X	X	X		X				FOR IN-HOUSE USE
			x	X	x	x		X	X	X	SELL SERVICE
x	x				X	X	X	X	x	X	SELL SOFTWARE 2.3.1, 2.3.3
		x		x		X	X	X	X	X	MARKETING ORIENTATION
X	X		X	X	X				X		RESEARCH ORIENTATION
M	C	B	B	C	U,C	I	I	I	I	I	TYPE OF COMPUTER 2.3.14
2	1	3	4	1	1	5	5	5	5	3	PROGRAMMING LANGUAGE 2.3.15

X = very much so x = not as much so

Large Textual Data Bases

A = the orientation is bibliographic.

B = the orientation is other than bibliographic.

a or b = large textual data bases are not the major focus of the system at this time.

Management Information

A = the system has host language capabilities as defined by the CODASYL committee.

Full Text Searching

x = the position of words in text has not been stored.

Telephone Network

STAIRS is on an internation IBM in-house network.

LEADER and SPIRES are on regional university networks.

NASIS relies on the federal telephone network.

The rest are on Tymshare's national network.

DIALOG also operates a national high-speed (480 cps) network.

Type of Computer

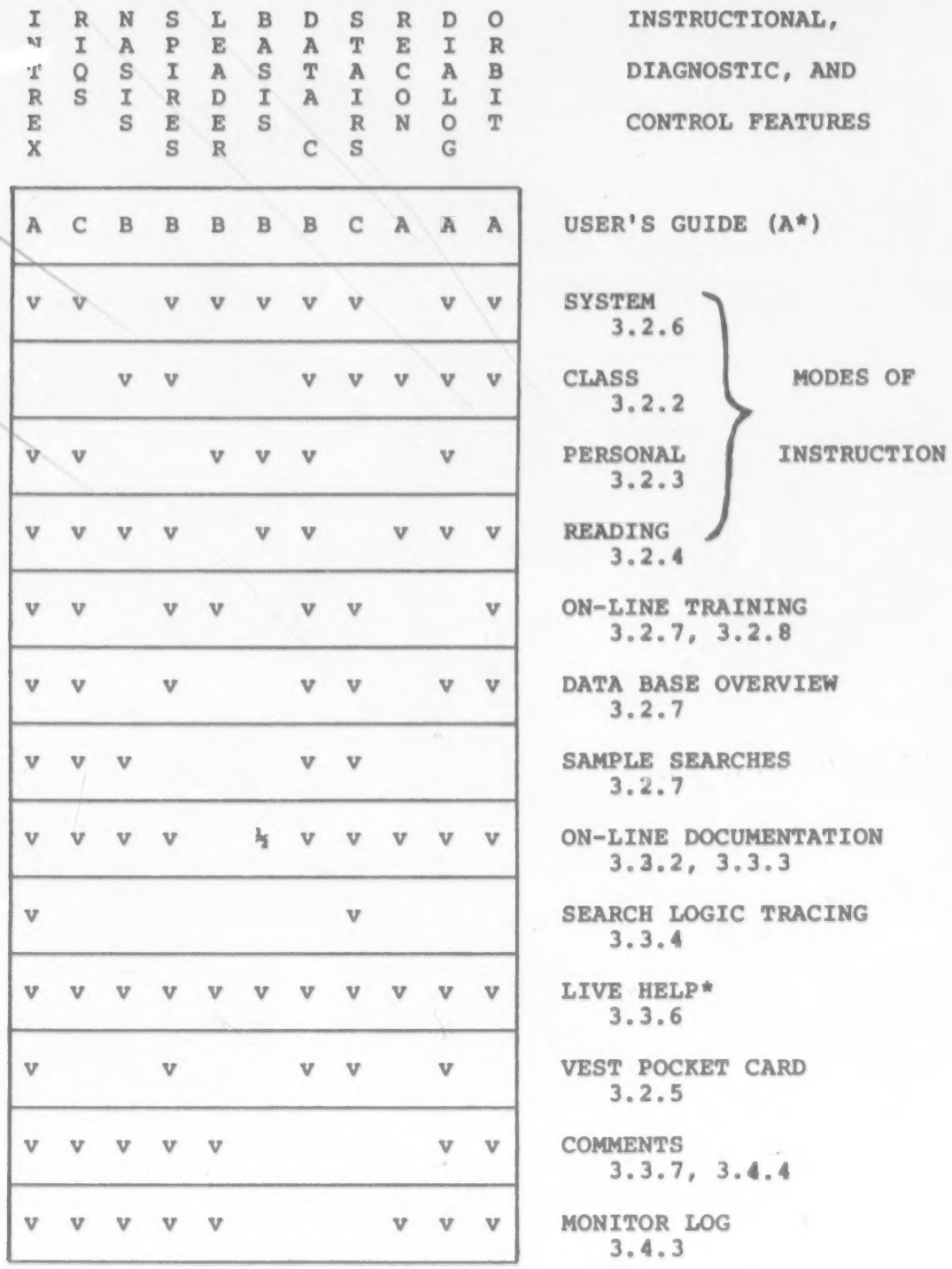
I = wide range of IBM 360's and 370's, B = IBM 360/67,

M = IBM 7094, C = CDC 6000 series, U = UNIVAC 1108

Programming Language (90-95% of the code)

1 = Fortran, 2 = extension to ALGOL, 3 = PL/1,

4 = PL/360, 5 = Assembly language.



INTRODUCTORY	QUICK REFERENCE	NASIS	SPECIAL FEATURES	LEADER	BASIC	DATA ACCESS	STATUS	RECORD	DIALOG	OBJECT	INSTRUCTIONAL, DIAGNOSTIC, AND CONTROL FEATURES
A	C	B	B	B	B	B	C	A	A	A	USER'S GUIDE (A*)
v	v		v	v	v	v	v		v	v	SYSTEM 3.2.6
		v	v			v	v	v	v	v	CLASS 3.2.2
v	v			v	v	v			v		PERSONAL 3.2.3
v	v	v	v		v	v		v	v	v	READING 3.2.4
v	v		v	v		v	v			v	ON-LINE TRAINING 3.2.7, 3.2.8
v	v		v			v	v		v	v	DATA BASE OVERVIEW 3.2.7
v	v	v				v	v				SAMPLE SEARCHES 3.2.7
v	v	v	v		½	v	v	v	v	v	ON-LINE DOCUMENTATION 3.3.2, 3.3.3
v							v				SEARCH LOGIC TRACING 3.3.4
v	v	v	v	v	v	v	v	v	v	v	LIVE HELP* 3.3.6
v			v			v	v		v		VEST POCKET CARD 3.2.5
v	v	v	v	v					v	v	COMMENTS 3.3.7, 3.4.4
v	v	v	v	v				v	v	v	MONITOR LOG 3.4.3

MODES OF INSTRUCTION

* = an essential feature

v = has been implemented

½ = currently being implemented

User's Guide

Each user needs a single document that in simple terms explains 1) how to use a terminal, 2) how to log onto and off of the retrieval system, 3) what commands are available and how they work, 4) what data bases are accessible and how they are structured, 5) how to develop a good search strategy, and 6) what to do when X happens.

A -- both complete and readable

B -- incomplete but readable or
complete but not readable

C -- neither complete nor readable

Modes of Instruction

The major approach(es) taken when educating the first-time user.

SYSTEM -- the system is programmed to help in the training

CLASS -- live or filmed courses are given periodically or upon request

PERSONAL--an expert sits down with the user at the terminal

READING-- the manual is read

On-line Training

Information concerning 1) terminal and typing problems, 2) the command repertoire, 3) characteristics of the data base, 4) hints for good searching, 5) common pitfalls and their remedies.

ORBIT -- The beginning user easily falls into this material.

DATA CENTRAL -- The beginning user easily falls into this material. When in verbose mode, system responses cue the user.

SPIRES -- The beginning user easily falls into this material. He can go through a sample dialog annotated with explanations.

RIQS -- The knowledgeable user can attach a tutor to the retrieval system. The tutor is the unique source of command and error explanations.

INTREX -- When in verbose mode, system responses cue the user.

LEADER -- When in verbose mode (the only mode), system responses cue the user.

Data Base Overview

Information concerning 1) the size of the data base, 2) the types of material in it, 3) the time span covered by the material, 4) the fields making up a record, 5) the searchable fields, and 6) special strategies to use or pitfalls to avoid with the data base.

Sample Searches

Prestored searches that teach by analogy. Unfortunately, most of the existing implementations assume the user knows the names of the prestored search strategies.

On-line Documentation

One page descriptions of all commands and error messages.

Search Logic Tracing

The searcher can request a detailed description of how his multi-part search request lead to the number of hits reported.

Live Help

Either a telephone number to call if desperate or a message command for requesting help from the on-line human consultant.

Vest Pocket Card

A durable folder containing command names and an explanation of how to get complete command descriptions.

Comments

The user can express his feelings or contribute insights to the system staff. Most useful in conjunction with a system log so that the staff can examine the user's pre-comment behavior.

Monitor Log

Typically a list of the commands entered by the user during his session.

INTREX -- Able to have two terminals see and control a single display (as well as traditional monitoring)

NASIS -- Individual sessions are not logged. Instead, every week command and data base field usage is summarized.

INTREX	RIQS	NASIS	SPIRES	LEADER	BASIS	DATAC	STAIRS	RECON	DIALOG	ORBIT	QUERY FORMULATION FEATURES
x			v	v,x		v,x	v	v	v	v	SUFFIX REMOVAL* 4.3.13-14, 4.4.5
	v	v			v	x	x	v	x	x	SEARCH FIELD CONTROL*
		v	v		v	v	v	v	v	v	DICTIONARY ACCESS* 4.4.7, 4.4.8
	v	v	v			v	v	v			RELATIONAL OPERATORS* 4.4.3
					v,x	x	x	v	v	v	SPELLING VARIATIONS 4.4.4
			v	v,x	v			v	v	v	RELATED TERM CAPABILITY 4.4.9
v	x					v,x	v	v,x	v	v	WORD PROXIMITY OPERATORS 4.3.16-17, 4.4.16-18
2	1	1	2		3	3	1	1	1	1	BOOLEAN OPERATORS* 4.5.1-4
v	v	x			x		x	x	x	x	REQUEST SETS* 4.5.7, 4.5.13
x			x	x		x	x				PHRASE DECOMPOSITION 4.3.12
		v	v		½		v	½	v	½	SEARCH PROFILES 4.5.14
v	v	v			v		v			v	SEQUENTIAL SEARCHING 4.5.15

* = an essential feature v = under user control

½ = currently being implemented x = assumed automatically

Suffix Removal

Most systems require the user to indicate suffix removal by entering the root followed by a truncation code. Automatic stemming in INTREX can be overriden by following the term with an exclamation point. DATA CENTRAL automatically deletes final s's.

Search Field Control

If the user does not specify which field he is searching, an X means all fields are assumed while a V means a single field (or a combined field like title and abstract) is assumed. The user can override the default by specifying a field name.

Dictionary Access

Alphabetic portions of the controlled vocabulary, index phrases, or word dictionary are displayed. Only DATA CENTRAL and SPIRES do not specify the number of documents associated with each term.

Relational Operators

The operators GREATER THAN, LESS THAN, and BETWEEN can be used with numeric fields. In STAIRS, the operators can only occur when performing a sequential search.

Spelling Variations

Words like 'avenue' and 'ave.', or 'January' and 'Jan.' are linked so that the use of any one incorporates the others.

Related Term Capability

Where terms in the vocabulary of a data base point to other terms, the user can see the other terms. LEADER has both automatic phrases and optional associated phrases. DIALOG, RECON, and LEADER number the terms so they can be incorporated by number.

Word Proximity Operators

DIALOG, STAIRS, and RIQS allow the user to specify that the terms are to occur in the same field without specifying the exact field. DIALOG, RECON, and DATA CENTRAL allow the user to specify how many words may separate two terms. All systems listed in the word proximity column permit testing for exact phrase matching. ORBIT and SPIRES provide for phrase matching via sequential search.

Boolean Operators

Generally OR is used to combine related terms, AND is used to intersect distinct concepts, and AND NOT is used to exclude previously displayed material. A 1 indicates that AND's are processed before OR's. A 2 indicates that the leftmost operator is processed first. A 3 indicates that OR's are processed before AND's, but, by using variants of the operators, AND's can be processed before OR's.

Request Sets

Each line of search request input is given a set number. Later search requests can incorporate earlier sets by mentioning the set number.

Phrase Decomposition

A natural language phrase is decomposed into significant words. Common words like 'the' and 'about' are deleted. INTREX assumes a logical AND between terms. SPIRES assumes either AND or OR depending upon the file definition. LEADER ranks retrieved phrases avoiding the problem.

Search Profiles

A user can develop a search strategy (or components of a search strategy), store the strategy, and rerun it many days later. DIALOG permits users to develop and store search strategies so that they can automatically be run against file updates.

Sequential Searching

In order to search fields that have not been inverted, each record in the set is examined one after the other. This is the only method for searching in RIQS.

INTREX	RIQS	NASIS	SPIRES	LEADER	BASIS	DATAC	STAIRS	RECON	DIALOG	ORBIT	RESULT MANIPULATION FFATURES
		v		v	v	v	v	v	v	v	SEARCH REVIEW* 4.5.9
v		v	v	v		v		v	v	v	PREDEFINED FORMATS* 5.2.6, 5.3.1
1	4	4	1		2	2	1	3	2	1	ON-LINE FORMATTING* 5.2.7, 5.3.2-5
v	v		v			v	v	v	v		RAPID SCAN 5.2.5
v				v		v	v	v			HIGHLIGHTING 5.2.9, 5.3.9
v		v		v	v	v	v	v	v	v	EXPANDING 5.3.10, 5.3.11
	½	½	½		½	v	v	v	v	½	SORTING 5.3.6
				v			v			v	RANKING 5.3.7
	v	v	v		v	v					COMPUTING 5.3.8
v					v				v		MICROFICHE 5.3.16
	v				v						DISPLAY OF GRAPHS 5.3.17
	v				v				v		STATISTICAL INTERFACE 5.3.18
v	v	v	v	v	v	v	v	v	v	v	OFF-LINE PRINTING* 5.3.13-15
	v	v		v	v		v	v	v	v	BATCH RETRIEVAL 2.3.7
						v	v		v		PHOTOCOMPOSITION
		v	v		½		v			½	DATA ACCESS PROTECTION 3.4.2

* = an essential feature v = has been implemented

½ = currently being implemented

Search Review

In order to review what has been done, a short summary is presented of the sets currently active, the number of documents each contains, and the request that caused the set to be created.

Predefined Formats

To allow the user to easily vary the way in which he views retrieved documents. Usually one format displays descriptive fields, another short citations, and a third displays total documents. With management information the predefined formats are likely to specify page layouts.

On-line Formatting

To allow the user to indicate the fields of each document he wishes to see. A 1 indicates that the fields can be put out in one order; a 2 indicates that they can be put out in any order; a 3 indicates that limited page organization is possible; a 4 indicates full scale on-line report formatting.

Rapid Scan

To quickly look over a little information about each document retrieved. The display should continue until the user interrupts it. DIALOG and LEADER simulate rapid scan by stacking commands. RIQS provides for nothing but rapid scan.

Highlighting

To find out what it was in the document that caused it to be retrieved. The INTREX feature does not highlight text but rather puts out only those field values that caused retrieval.

Expanding

To have a complete document displayed after seeing a portion of it. This is usually done either by entering a sequence number or document number.

Sorting

To sort the documents according to the values of a designated field or fields.

Ranking

To rank the documents so that the ones containing the greatest number of search terms are listed first. STAIRS provides for five different ranking options. LEADER automatically ranks output.

Computing

BASIS and RIQS allow the user to define variables as functions of field values. The calculated values may be displayed or operated upon by statistical subroutines. In SPIRES, the average command combines calculating the value and displaying the result.

Microfiche

The terminal controls access and display of the microfiche. In DIALOG the feature is not used.

Display of Graphs

Graphics terminals are used by RIQS and BASIS for on-line plotting of fields from the document set retrieved. In neither case is it possible to superimpose graphs from different sets.

Statistical Interface

The records retrieved by searching can be passed to an on-line statistical package for analysis.

Off-line Printing

The documents are printed and (usually) are mailed to the user. The user must divert the request to the batch system with both RIQS and SPIRES.

Batch Retrieval

There is communication from the on-line system to the batch system with both NASIS and LEADER. In the other cases, the batch retrieval command language is similar to the on-line language.

Photocomposition

In close association with the retrieval system is an off-line photocomposition system for publishing data base documents.

Data Access Protection

Access to records within the data base or fields within records can be restricted to a subset of all users.